Social Science Library
JUN 2 - 19

Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

CONTENTS

Acquisition, Extinction, and Recovery Functions in Retroactive Inhibition: G. E. BRIGGS .. 285
Studies of Distributed Practice: XII. Retention following Varying Degrees of Original Learning: B. J. UNDERWOOD 294
The Principle of Selective Association of Drive Stimuli: C. B. WOODBURY AND D. H. WILDER 301
Transmitted Information and the Length of Rating Scales: A. W. BENDIG 303
Tracking and Frequency of Target Intermittence: W. F. BATTIG, L. W. GREGG, E. H. NAGEL, A. M. SMALL, JR., AND W. J. BROGDEN 309
Response Time Patterns Associated with Various Display-Control Relationships: W. D. GARVEY AND W. B. KNOWLES 315
Shape Discrimination as a Function of the Angular Orientation of the Stimuli: M. D. ARNOULT 323
Effect of Stimulus Area and Intensity upon the Light-Adapted Electroretinogram: J. C. ARMINGTON AND F. C. THIEDE 329
Perception of the Relative Distance Position of Objects as a Function of Other Objects in the Field: W. C. GOGEL 335
The Effects of Punishment upon Syllable Recognition Thresholds: W. LYSAK 343
A Test for Conditioned Inhibition in Motor Learning: J. A. STARKWEATHER AND C. P. DUNCAN 351
Warm-Up in Massed and Distributed Pursuit Rotor Performance: A. M. BARCH 357
Two Processes in Perceptual Learning: F. P. KILPATRICK 362
The Effect of Kinesthetic, Verbal, and Visual Cues on the Acquisition of a Lever-Positioning Skill: W. F. BATTIG 371

American Psychological Association

Vol. 47 No. 5 May 1954

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1954 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas.

Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Copyright, 1954, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 47, NO. 5 MAY, 1954

ACQUISITION, EXTINCTION, AND RECOVERY FUNCTIONS IN RETROACTIVE INHIBITION[1]

GEORGE E. BRIGGS[2]

Northwestern University

In his discussion of the reproductive form of retroactive inhibition, McGeoch states that: "The character of this inhibition is most readily understood in terms of competition between systems of response. . . . On this view, a response is inhibited when another response takes the place of it. . . . Inhibition is not thought of as a separate process, but as a function of competition among responses, with a resultant momentary dominance, at least, of one response over another. Responses thus inhibited are not necessarily lost from the subject's repertoire, but are kept by other responses from appearing. It is easy to understand, then, why a once-learned response may not appear under one condition and may appear under another" (5, pp. 494–495).

It is with the reproductive inhibition situation that the present study is concerned. Here an effort was made to measure the course of the inhibitory processes during the learning of the original and interpolated materials. Further measurement was made to determine the relative strengths of the competing response systems as a function of time following the learning sessions. The method used to determine these systems was similar to a technique utilized by Underwood (9) who describes it as modified free recall. Essentially, modified free recall is a noncorrection test for retention wherein the stimulus items used in the original and interpolated learning sessions are presented alone. The *S* is requested to recall either of the previously learned responses, if possible. If this cannot be done, he is free to give any response. Therefore, this technique would be expected to more accurately define the "competition between systems of response" indicated by McGeoch as a fundamental basis for retroactive inhibition.

METHOD

Subjects and materials.—The basic design of this experiment followed the reproductive inhibition paradigm with 125 *S*s in the experimental group and 25 *S*s in the control group. All *S*s

[1] A condensation of a doctoral dissertation which is on file in the Library of the University of Wisconsin. The author expresses appreciation to Professors W. J. Brogden, D. A. Grant, and H. F. Harlow under whose direction the dissertation was prepared. The research was supported in part by a grant to W. J. Brogden from the Research Committee of the Graduate School.

[2] Now a postdoctoral fellow of the National Science Foundation.

were students from elementary psychology classes, and none had served previously in an experiment on retroactive inhibition. For the experimental group the list learned in OL and in IL was composed of 12 paired adjectives with the stimulus words the same in both lists. The 12 stimulus words were selected from a sample of 50 common two-syllable adjectives which had been presented to a total of 100 elementary psychology students in a free word-association situation. Selection of the 12 words was on the basis of minimum variability of response.

General procedure and apparatus.—The general procedure on the first day was to give each *S* a short practice session of three trials on a list of five paired-associate nonsense syllables. Each *S* then learned the original list to a criterion of one perfect trial. Twenty-four hours later the interfering list was learned, by the experimental group, to the same criterion. Subgroups of 25 *S*s each relearned the original list at intervals of 4 min., 6, 24, 48, or 72 hr. following IL. The control group followed the same general procedure except that IL was omitted on the second day. Subgroups of 5 *S*s were tested at the same intervals of the retention period as were used with the experimental group.

In order to facilitate conduct of the experiment, two *S*s learned the lists simultaneously. This was achieved by exposing the lists on two milk glass screens, one for each *S*, and requesting *S* to write his anticipations on a recording apparatus known commercially as a Tel-Autograph Transceiver. The responses appeared, in *S*'s own handwriting, both on the transceiver and on a receiver located in an adjacent room from which *E* projected the words. The two *S*s were separated by a screen, and headphones were available so that *E* could communicate with each *S* separately. The *S*s' and the *E*'s rooms were soundproof, and therefore the exposure, recording, and communicating systems enabled each *S* to work independently.

Two automatic projectors were used to expose the words. Each stimulus word and each stimulus-response pair were photographed on 35-mm. film and the negatives were mounted in 2 × 2-in. metal binders. Four orders of presentation were then arranged. Exposure of the slides was entirely automatic. A three-unit electronic timer was set to cycle continuously through two phases. Two units controlled a 4.5-sec. phase during which time a slide was exposed to *S*'s view. Termination of this phase activated a third unit of the timer, which in turn set the slide changer mechanism into operation. This third unit timed the .5-sec. interval required by the projectors to effect a change of slides. Termination of the third unit reactivated the two-unit phase, and so on. The intertrial interval was set at 10 sec.

Modified free recall and treatments.—Immediately after the short practice session with nonsense syllables, each *S* was presented with the 12 stimulus words from the OL and IL lists. He was instructed to give the first word to come to mind upon seeing each of the adjectives. By this free association test it was possible to determine the original responses which *S* brought into the situation with him, the R_0 class of response. At some predetermined point (see below) in either OL or IL or both, *S* was interrupted in his learning and given a modified free recall (MFR) test. The following instructions were given for MFR during OL: "I am now going to present the stimulus words alone. Just as soon as you see an adjective come on the screen, write down the first word that comes to mind. If you can recall the response that has been assigned to each stimulus word, put that down, of course. If, however, you are unable to recall that response, then put down whatever word you think of upon seeing each of these terms."

The instructions given for MFR during IL and at the retention intervals following IL were expanded slightly to request that either of the two responses assigned to each stimulus word be given, if possible. The responses assigned to the first list are designated as R_1 and those in the interpolated list are labeled R_2. Any response given during MFR which was incorrect, in either list, for the particular stimulus word to which it was given, will be termed an R_3 class of response. Finally, any responses which were from neither of the two lists are classified with the original responses given by *S*, the R_0 class.

Four points of interruption were selected for both OL and IL. During OL an *S* might have been given an MRF test after the trial on which he had correctly anticipated 3 responses ($\frac{1}{4}$ criterial level), 6 responses (the $\frac{1}{2}$ criterial level), 9 responses (the $\frac{3}{4}$ criterial level), or 12 responses (the $\frac{4}{4}$ criterial level). As a control treatment within the experimental group, some *S*s were not interrupted at all during OL or IL or both. This is designated as the zero treatment. No *S* was interrupted more than once during either OL or more than once during IL, and after interruption, all *S*s went on to the $\frac{4}{4}$ criterion of learning if interruption occurred prior to that level. These MFR tests give an index to determine (*a*) the acquisition function for the R_1 responses and the simultaneous extinction function for the R_0 responses during OL, (*b*) the acquisition function for the R_2 response class and the simultaneous extinction curve for the R_1 class during IL, and (*c*) the changes in strength of the R_3 class during both OL and IL.

An MFR test was given to *S* just prior to the traditional recall and relearning of the original list after one of the five retention intervals (4 min., 6, 24, 48, or 72 hr.). This permits a determination of (*a*) the retention function for the R_1

response class, (*b*) the retention, or as they will be termed, the recovery functions for the R_0 and R_1 classes, and (*c*) the presence of R_3 responses over the 72-hr. period following the IL.

With the five possible treatments during OL and the five during IL (0, ¼, ½, ¾, 1) and the five retention intervals (4 min., 6, 24, 48, 72 hr.), there were 5 × 5 × 5 or 125 unique combinations of the experimental treatments. In order to complete one replicate, 125 *S*s were used with one *S* per treatment combination. This design breaks the experimental sequence of OL—IL—RL into five subgroups for each of the three sessions. Each subgroup contains 25 *S*s. This design for the experimental group gives no information on the relative frequency of response for the R_0, R_1, and R_3 classes just prior to IL. The IL session occurred 24 hr. after OL, and thus it was necessary to measure the changes which occurred in the three classes over the 24-hr. intersession interval.

Therefore, the 25 control group *S*s were given an MFR test 24 hr. after OL. They were then dismissed until time for recall and relearning. Subgroups of 5 *S*s were given the same treatment during OL, and each returned for recall and relearning at one of the same retention intervals as for the experimental group. As was the case for the experimental group, each control *S* received a unique treatment combination in OL and RL. No control *S* was interrupted for MFR more than once in OL nor was he tested for retention at more than one retention interval.

In summary, four of the five treatments (¼, ½, ¾, and 1) in both OL and IL and the five retention intervals (4 min., 6, 24, 48, and 72 hr.) gave *S* an opportunity to respond in an MFR test. There were 25 experimental group *S*s in each of these subgroups, and thus acquisition, extinction, retention, and recovery functions were each determined by at least four points with 12 × 25 or 300 possible responses at each point. The zero criterial level in IL was determined by the 25 *S*s of the control group.

Results

The results of the modified free recall (MFR) tests are presented in Table 1 and in Fig. 1. Table 1 contains the mean number of responses for each response class throughout the learning of the two lists and over the 72-hr. retention period following IL. Standard error values are also listed, where appropriate, for the means, and both means and their associated standard errors are based on an *N* of 25. In Fig. 1 each score is expressed relative

TABLE 1

Mean Frequency for Response Classes R_0, R_1, R_2, and R_3 on MFR Tests

Condition	Response Class	Criterial Levels									
		0		¼		½		¾		1	
		M	σ_M	*M*	σ_M	*M*	σ_M	*M*	σ_M	*M*	σ_M
OL	R_0	12.00	—	5.88	.54	3.72	.46	1.60	.31	0.96	—
	R_1	0.00	—	4.24	.43	6.24	.41	9.36	.32	10.68	.36
	R_2	—	—	1.88	.32	2.04	.41	1.04	.21	0.36	—
IL	R_0	2.16	.41	2.72	.36	1.76	.38	0.84	—	0.40	—
	R_1	9.04	.45	5.32	.48	4.44	.54	3.52	.51	2.80	.45
	R_2	0.00	—	2.24	.33	4.40	.48	6.32	.52	7.96	.50
	R_3	0.80	—	1.72	.21	1.40	.19	1.32	.23	0.84	—
Condition	Response Class	Retention Intervals									
		4 min.		6 hr.		24 hr.		48 hr.		72 hr.	
		M	σ_M	*M*	σ_M	*M*	σ_M	*M*	σ_M	*M*	σ_M
RL	R_0	0.36	—	1.40	.27	0.84	—	1.80	.34	2.12	.44
	R_1	2.56	.48	4.36	.37	5.32	.44	4.96	.45	5.04	.47
	R_2	8.20	.61	5.20	.41	4.80	.42	4.08	.46	3.80	.51
	R_3	0.88	—	1.04	.23	1.04	.20	1.16	.22	1.04	.21

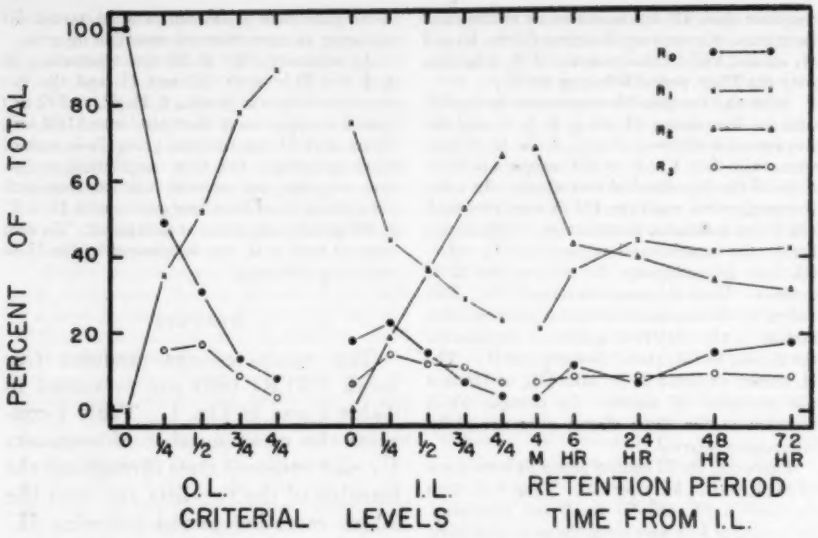


FIG. 1. Percentage of response for the four response classes during OL and IL and over the 72-hr. retention period

to a total of 300 responses possible at each ordinal point.

The base of the graph in Fig. 1 is divided into the three main sessions of the experiment: original learning, interpolated learning, and retention period. Within OL and IL further subdivision is made for each point of interruption for MFR, the criterial levels. The zero point for OL was assumed to give 100% of the R_0 class and 0% of the R_1 class. The zero level for IL contains the responses from the control group *S*s who returned 24 hr. after OL for an MFR test. This is the only place that control group scores appear in either Fig. 1 or Table 1.

Acquisition and extinction.—In both OL and IL the dominant processes are acquisition and extinction. From visual inspection, the acquisition function for the R_1 class appears to be similar to the acquisition curve for the R_2 class in IL in that both are negatively accelerated functions with positive slope. The most striking difference between the two curves is that the R_2 function in IL does not reach the maximum level attained by the R_1 function in OL. There also appears to be considerable similarity between the R_0 extinction function in OL and the R_1 extinction function in IL in that both are negatively accelerated with a negative slope. Again, however, the IL function does not reach the final level attained by the OL curve. All four functions display components of linearity over major portions of their total extent.

During OL the R_1 acquisition function reveals a relationship between criterial level, as shown by the ordinal values of the graph, and the relative level of learning which was determined by the MFR tests. Early in the OL session, the ¼ criterial level, *S* gave considerably more R_1 responses during MFR than would be expected at that criterion.[3] The mean value of 4.24

[3] The criterial levels of ¼, ½, and ¾ do not necessarily indicate that all *S*s were interrupted

from Table 1 is 1.24 units greater than the observed value of three correct anticipations on the $\frac{1}{4}$ criterial trial. At the $\frac{3}{4}$ and $\frac{2}{4}$ criterial levels *S* still gave more R_1 responses during MFR, 6.24 and 9.36 respectively, than would be expected on the basis of performance on the criterial trial. However, the differences between expected and obtained values at these points are far less than that difference at the $\frac{1}{4}$ criterial level. Also, these differences are small compared to the associated standard errors of .41 and .32. After the $\frac{1}{4}$ criterial level had been reached, however, *S* gave a mean of only 10.68 R_1 responses during MFR. This is significantly less than the 12 correct responses given on the criterial trial just prior to MFR. The results of this study for the $\frac{3}{4}$ and $\frac{1}{4}$ criterial levels are in agreement with Ward (10) where the same general findings held for the $\frac{7}{12}$ and $\frac{1}{12}$ criterial levels in his study of reminiscence in serial learning.

This relationship does not hold for the IL session of this study. There, level of mastery of the R_2 responses, as measured by the MFR tests, was always less than that indicated by the performance on the criterial trials. Further, the difference between expected and obtained levels increases as learning progresses.

Retention and recovery.—The third section of Fig. 1 contains a retention function for the R_2 response class beginning at 4 min. after IL and extending over a 72-hr. period. This curve for retention follows the usual course of such functions and is quite similar to that retention function demonstrated by Underwood (9) in a comparable experimental situation. The companion process indicated here is the recovery of the R_1 class over the retention period. An analysis of variance was applied to the data to test for significant differences between R_1 and R_2 over the five retention intervals. For this analysis the data were transformed to arc sin (per cent)$^{\frac{1}{2}}$ (8). The results are presented in Table 2.

TABLE 2

RESULTS OF ANALYSIS OF VARIANCE ON THE FREQUENCY OF RESPONSE FOR R_1 AND R_2 OVER THE FIVE RETENTION INTERVALS

Source	*df*	Mean Square	*F*
1. Between *Ss*	124		
a. Groups	4	239.336	3.93*
b. *Ss* within groups	120	60.831	
2. $R_1 - R_2$	1	1753.664	5.18*
3. Within *Ss*	124		
a. $R_1 - R_2$ × Groups	4	4313.246	12.73*
b. $R_1 - R_2 \times Ss$ within groups	120	338.695	

* Significant at or beyond the .05 level of confidence.

In the analysis, the *F* value for Groups (1*a*) was obtained by testing against *Ss* within groups (1*b*), and the *F* values for $R_1 - R_2$ (2) and $R_1 - R_2 \times$ Groups (3*a*) were obtained by testing each against $R_1 - R_2 \times Ss$ within groups (3*b*). The *F* ratios in Table 2 indicate that the following are significant sources of variation: Retention Intervals, R_1 vs. R_2, and the $R_1 - R_2 \times$ Retention Intervals interaction. This latter significant source justifies the use of Duncan's range test (2) to test for significant differences between the ten means defining the R_1 and R_2 functions in the retention period (see Table 1).

From the range test the following points on R_1 and R_2 were found to be significant: the initial values for both R_1 and R_2, 2.56 and 8.20 respectively,

for MFR after the trial on which they had correctly anticipated exactly 3, 6, or 9 responses, respectively. It did happen that during OL all *Ss* had not exceeded the criteria before MFR, but during IL some *Ss* did exceed criteria before test was possible. The actual median number of correct responses on the criterial trials during IL was 3.46, 6.54, and 9.28.

were significantly different from the other eight means, and likewise, the two final values for R_2, 4.08 and 3.80, differ significantly from the other means. Further, 2.56 differs significantly from 8.20, but no significant difference was noted between 4.08 and 3.80. The .05 level of confidence was used for determining significance.

From these analyses, then, the retention function for R_2 indicates a significant decrease in response strength between 4 min. and 6 hr. with further significant, but more gradual, decrease over the remainder of the retention interval. The recovery function for the R_1 class increases significantly between 4 min. and 6 hr. The maximum effective level of R_1 response strength in the retention period is reached between 6 and 24 hr., and this level is maintained over the remainder of the 72-hr. period. This maximum level for the R_1 class occurs after the R_2 function has decreased to approximately 40% of total. It may therefore be concluded that the R_2 retention function represents a dominant process or response system.

The recovery function here differs from that found by Underwood (**9**) in that the present curve rises to a higher maximum level and has a greater slope. Also, the recovery function in the Underwood report reaches maximum effective level when the retention function is near the 50% level.

R_0 and R_3 functions.—Following the extinction of R_0 during OL, Fig. 1 shows that further extinction takes place during the IL session. From visual inspection, the frequency of response for this class appears to increase again over the retention period. However, an analysis of variance similar to that summarized in Table 2 failed to support this observation as significant statistically. From this analysis of variance only the retention intervals variable was indicated to be a source of significant variation. Therefore, the most that can be said for the R_0 function here is that if it represents a recovery process, such recovery is progressing at a very slow rate.

The function for the R_0 class during the retention period is quite similar to the "other" curve presented by Underwood (**9**, p. 432–433) if that latter function is corrected to the criteria defining the R_0 class of the present study. If so corrected, Underwood's results for R_0 are .40, 12.90, 7.10, and 19.20% for the retention intervals 1 min., 5, 24, and 48 hr., respectively. The comparable results from the present study are 3.00, 11.67, 7.00, and 15.00% for the intervals 4 min., 6, 24, and 48 hr., respectively. Finally, the R_3 function shows its greatest strength during OL and IL with maxima during levels of partial mastery. Over the retention intervals, R_3 remains at a low and constant level.

Changes in response strength between OL and IL.—Figure 1 clearly shows that changes in response strength have occurred between the $\frac{4}{4}$ level of OL and the zero level of IL. An analysis of variance similar to that summarized in Table 2 was performed on the data, and as was the case in Table 2, all the comparisons were significant at or beyond the .05 level of confidence. These comparisons were between $\frac{4}{4}$ OL and zero level of IL, between the three functions R_0, R_1, and R_3, and the interaction of the functions and criterial levels. This latter significance justified the Duncan range test (**2**), which, when applied to the six mean values involved, showed that only the R_0 and R_1 functions changed significantly over the 24-hr. intersession interval.

Effects of MFR tests.—In a single over-all analysis of variance it was possible to (*a*) determine whether or not interruption for MFR during OL and IL had any significant effect on such RI as was found in this study, and (*b*) determine whether or not there were significant changes in RI as measured at the five retention intervals covering the 72-hr. period of time from IL. By a simple between-within

analysis of variance, the recall scores on the first RL trial for the control group showed no significant differences for the five retention intervals. Therefore these 25 scores were pooled to achieve a more reliable estimate of the mean recall score for the control procedure. Relative RI scores were then computed for each experimental group *S* and these were transformed to arc sin (per cent)$^{\frac{1}{2}}$ (**8**). The over-all analysis of variance tested the three factors of interruption for MFR during OL, interruption during IL and retention intervals, and their first order interactions against the second order interaction.

No statistical significance was indicated for any comparison. Thus there is no indication that the technique of MFR introduced a significant effect into OL or IL as indicated by measures of relative RI. Further, there appears to be no differential amount of RI displayed at the five retention intervals over a total of 72 hr. following IL. This latter finding is in agreement with a report by McGeoch (**4**) where RI was relatively constant when measured over a period of 144 hr.

Positive transfer from OL to IL.—The mean number of trials required to reach the $\frac{1}{4}$ criterion in OL was 11.03, and in IL it was 9.26. An analysis of variance similar to that presented in Table 2 indicated that the difference here was statistically significant at the .05 level of confidence.

Acquisition vs. extinction processes.—From Fig. 1 it would appear that the acquisition functions and the extinction functions in OL and IL have similar slopes. Therefore, Alexander's test for trend (**1**) was applied to the data defining these functions. The two acquisition functions formed one group and the two extinction functions formed the other group. This analysis of variance showed that whereas there were significant differences within groups, there was no significant difference between groups in either slope or mean magnitude. This finding suggests that Guthrie's hypothesis (**3**), that extinction represents the acquisition of an antagonistic response, may be applicable to the reproductive inhibition paradigm.

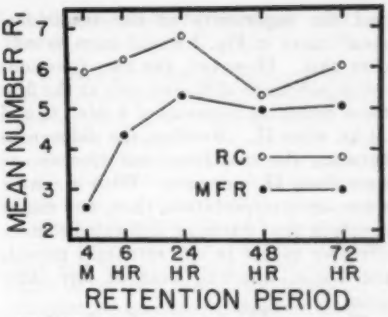


FIG. 2. Mean frequency of response during traditional recall (RC) and during modified free recall (MFR) at the five retention intervals

MFR vs. traditional recall, and relearning.—Figure 2 presents two recall functions for the five retention intervals of this experiment. The lower curve represents mean number of R_1 responses elicited during the MFR tests at each of the intervals. The upper function, RC, was determined by the frequency of R_1 responses correctly anticipated on the first relearning trial immediately after the MFR test. This latter function was determined under what has been termed traditional recall. Analysis revealed that the two functions differ significantly at the 4-min., 6-, and 24-hr. retention intervals. No statistically significant differences were detected at the 48- or 72-hr. intervals.

Finally, the relearning of the original list was carried out to the $\frac{1}{4}$ criterion. Analysis for the number of trials required to reach that criterion did not reveal any significant differences between the five retention intervals.

DISCUSSION

The differences between traditional and modified free recall have been pointed out. Each *S* had MFR just prior to what has been termed the traditional recall test. This temporal arrangement means that MFR could serve as a warm-up trial,

and the superiority of the traditional recall curve in Fig. 2 would seem to indicate this. However, the two functions are significantly different only at the first three retention intervals of 4 min., 6, and 24 hr. after IL. Further, the differences between the two functions decrease as time from IL increases. With a simple warm-up interpretation, then, one might conclude that warm-up is decidedly more effective earlier in the retention period, the 4-min. interval, than at any later time.

The retention function for the R_2 responses in Fig. 1 indicates a more logical mechanism. Early in the retention period the R_2 function is at a maximum value of 68.33%. At this same interval, the discrepancy is greatest between MFR and RC in Fig. 2. As the retention function for R_2 decreases over the 72-hr. period, so do the differences between traditional and modified free recall. The R_2 responses as a class are clearly stronger than the R_1 class early in the retention period, but this does not mean that the R_1 response is not available for recall. MFR permits *S* to give either OL or IL responses, but he is requested to give only the *first* of the two responses that he can recall. Due to the strong competition from R_2 responses, the R_1 class is held to an effective level of 21.33% during the 4-min. MFR test. Given an opportunity, as in traditional recall, *S* could recall not only those R_1 responses strong enough to appear in MFR, but also any R_1 responses of sufficient strength to be recalled but not given during MFR because of a prior recall of an R_2 response.

As response strength of the R_2 class decreases, so would the effective competition between R_2 and R_1 responses be expected to decrease. It logically follows that more and more of the R_1 responses would then appear during MFR as compared to an earlier retention interval. This would lead to less difference between MFR and traditional recall functions as time from IL increases. The data are in agreement with this hypothesis, and therefore warm-up, although it may be present to some extent, does not account for the observed differences as logically as does the factor of the effective level of R_2 during the retention period. This situation, then, demonstrates the validity of McGeoch's statements in that a once learned response is not necessarily lost from *S*'s repertoire. Instead, due to the momentary dominance (of the R_2 class early in the retention period), the once learned response (the R_1 class) is held to a low level during one recall situation (MFR) but not during another situation (traditional recall).

The matter of terminology should be made explicit here. The present report has used the term "extinction" to describe the R_0 function during OL and the R_1 function during IL. The justification for this application is the communality between the experimental conditions facing *S* in this verbal learning situation and those present during a formal conditioning study. In both cases a progressive decrement in the frequency of response occurs when a previously learned response is no longer appropriate to the stimulating conditions.

It should be pointed out that the extinction function for the R_1 class does not necessarily indicate the course of unlearning during IL. The present data offer no basis for evaluating unlearning as that process has been defined (6, 7). Therefore, possible differentiation of the two processes, extinction and unlearning, must await future research.

In the preceding sections, comparisons between the results of this experiment and those from the Underwood report (9) were made. It was pointed out that the recovery function for the R_0 responses was quite similar to Underwood's function for the "other" responses if that latter function was corrected to the criteria defining the R_0 class. There was a decided difference, however, in the forms of the recovery functions for those responses learned during OL. The R_1 function in this study has a steeper slope and reaches a greater maximum than the comparable function from Underwood.

It is probable that the parameter of number of trials taken to learn in OL vs. those required in IL is an important factor in this difference. In the present study, there was highly significant positive transfer from OL to IL. From Underwood's data, however, there is no such degree of positive transfer indicated. There is positive transfer from OL to IL in that study, but the difference between trials to learn in OL and in IL does not appear to approach statistical signifi-

cance. In that experiment, therefore, the R_1 responses as a class would not be expected to have as great a superiority in response strength over the R_2 class as that indicated in the present study. This may account, in large measure, for the differences in the form of the two recovery functions.

Finally, it is important to note the differences between traditional and MFR techniques as measures of learning and retention. In traditional recall, major emphasis is placed upon one response class, the OL responses in the retroactive inhibition paradigm or IL responses in the proactive inhibition situation. With MFR, however, *S* is placed under no such restrictions because any response can be "correct" in an MFR test. In traditional recall, only incidental information (i.e., intrusions) is obtained for response classes other than the emphasized class, while in MFR *S* is actually encouraged to give an "intrusion" when a previously learned response cannot be recalled. Further, he is then not formally corrected for his "error." In MFR the effective level of several competing response classes or systems is given opportunity of expression, and therefore this technique provides more information on the dynamics of the transfer situation than can be provided by the traditional method of recall.

Summary

In an experiment on retroactive inhibition, 125 experimental group *S*s learned two lists of 12 paired adjectives. A 24-hr. period separated OL and IL, and recall and relearning of the OL list occurred at one of five retention intervals over a 72-hr. period. Each *S* was given modified free recall (MFR) at one of four criterial levels during OL and/or during IL, and each had an MFR test just prior to the traditional recall of the original list responses. Control group *S*s learned in OL with the same treatment as that for the experimental *S*, then returned for MFR 24 hr. after OL. Subsequent recall of OL responses was then carried out at one of the same retention intervals.

The MFR tests permitted determination of the following functions: (*a*) acquisition of the OL responses (R_1) and simultaneous extinction of the original responses (R_0) which *S* made to the stimulus words, (*b*) acquisition of the IL responses (R_2) and simultaneous extinction of the R_0 and R_1 classes, (*c*) retention of the R_2 responses and recovery of the R_0 and R_1 classes over the retention period, and (*d*) the behavior of an R_3 class which contains incorrect responses. The results show:

1. Recovery of the R_1 responses during the retention period eventually reached a level of approximately 40% of total as retention of the R_2 class decreased. The recovery function remained rather constant after attaining this level, whereas the retention function continued to decrease over the entire 72-hr. period. Only a suggestion of recovery for the R_0 class was noted.
2. As shown by MFR, acquisition and extinction were both carried to a greater extent in OL than in IL even though the $\frac{1}{1}$ criterion was common to both sessions. However, acquisition, as a process common to both OL and IL, did not differ in form (mean magnitude and slope) from the extinction process. This was offered as support for the hypothesis that extinction represents the acquisition of an antagonistic response.
3. A difference between the frequency of R_1 recalls during MFR and during traditional recall was related to the dominance of the R_2 responses early in the retention period and their continued effect over the entire period.

References

1. Alexander, H. W. A general test for trend. *Psychol. Bull.*, 1946, **43**, 533–557.
2. Duncan, R. B. A significance test for differences between ranked treatment in an analysis of variance. *Virginia J. Sci.*, 1951, **2** (new series), 171–189.
3. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.
4. McGeoch, J. A. Studies in retroactive inhibition: I. The temporal course of the inhibitory effects of interpolated learning. *J. gen. Psychol.*, 1933, **9**, 24–42.
5. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
6. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 157–173.
7. Melton, A. W., & Von Lackum, W. J. Retroactive inhibition in retention: evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, **54**, 157–173.
8. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Iowa State Coll. Press, 1946.
9. Underwood, B. J. 'Spontaneous recovery' of verbal associations. *J. exp. Psychol.*, 1948, **38**, 429–439.
10. Ward, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, **49**, No. 4 (Whole No. 220).

(Received August 6, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

STUDIES OF DISTRIBUTED PRACTICE: XII. RETENTION FOLLOWING VARYING DEGREES OF ORIGINAL LEARNING

BENTON J. UNDERWOOD [1]

Northwestern University [2]

In a previous study (6) we summarized fragmentary evidence which suggested that whether or not distributed practice of rote-learned materials influenced retention depended upon the degree of original learning. More specifically, the evidence indicated that items which were reinforced many times were better recalled following massed practice than following distributed practice; the reverse was true for items given only a small number of reinforcements during original learning. The evidence for these principles was only suggestive, but it was shown that certain contradictions in empirical findings, especially the contradiction between Hovland's (2) and our own work (5, 7), could be resolved if these principles were sound. The experiments in the present report are concerned primarily with this problem. We have manipulated the degree of original learning under massing and under distribution for both serial lists of nonsense syllables and paired-adjective lists. Observations are then made of retention as a joint function of degree of learning and intertrial interval during original learning.

A secondary purpose of the present experiments was to examine the effects of change of context between learning and relearning. In all of our previous studies relearning has been by massed practice regardless of whether original learning took place under massed or distributed practice. In these previous studies with serial lists, we have consistently found relearning to be as good (or better) following massed practice as following distributed practice. It is conceivable that such findings could result from the fact that both original learning and relearning were under massing in the one case, whereas in the case of distributed learning, relearning took place with the context changed. Such a change in context might influence relearning rate, although it should not influence recall. In the present experiments a factorial design has been used in which half of the *Ss* who learned under distributed practice relearned under massed, and the other half relearned with distributed practice. Likewise, half of the *Ss* who learned under massed practice relearned with massed practice, and half relearned with distributed practice.

METHOD

General.—Four groups of *Ss* were used, two groups learning serial nonsense syllable lists and two learning paired-adjective lists. For each kind of material one group went through three experimental conditions under massed practice and the other group under distributed practice. The three experimental conditions for each group differed only in terms of degree of original learning required before introduction of the 24-hr. retention interval.

The serial lists were made up of 14 syllables of low intralist similarity. These lists have been reproduced elsewhere (5). The *S* actually learned only 13 syllables since the first one was used as an anticipatory cue. For the low degree of learning, the list was presented until *S* achieved

[1] Jack Richardson supervised the running of the subjects and is largely responsible for the statistical analysis.

[2] This work was done under Contract N7onr-45008, Project NR 154-057, between Northwestern University and the Office of Naval Research.

seven correct responses on a single trial. For the moderate degree of learning, the list was presented until one perfect recitation was given. For the high degree of learning, the list was presented until one perfect trial was achieved, and then seven additional trials were added. For the massed group 2 sec. elapsed between each trial and for the distributed group, 30 sec.

The paired-adjective lists consisted of 10 pairs. The three experimental lists used had low intralist similarity and were known not to differ significantly in difficulty. These lists also have been given in an earlier report (1). The low degree of learning for these paired-associate lists consisted of five out of ten responses correct on a single trial; the moderate degree of learning was one perfect recitation; and the high degree, one perfect recitation plus seven trials. Massed practice consisted of 4 sec. between trials, with distributed practice having 30 sec. between trials.

Detailed procedure.—All groups were given one practice day to introduce them to the conditions of the experiment and to check on equality of the learning ability of the two groups learning paired associates and the two groups learning serial lists. On this practice day all *S*s learned a list to one perfect trial by massing. After a 5-min. rest they were given eight additional trials with 30-sec. rests between each trial. During these 30-sec. rests, and during all such rests on experimental days for the distributed group, symbol cancellation (4) was used to fill the interval.

There were 36 *S*s in each of the two groups that learned serial lists. For each group the design followed our previous studies in which both conditions and lists are independently counterbalanced. Each of the groups learning paired-adjective lists consisted of 24 *S*s. Lists and practice are confounded in the design, but lists, as indicated earlier, are known to be an insignificant source of variance. Conditions (three degrees of learning) were completely counterbalanced.

Relearning of all lists took place after 24 hr. For half of the *S*s relearning was by massed practice and for the other half relearning was by distributed practice. The sessions for each *S* came on five consecutive days. Relearning of a previous list and the learning of a new one took place during a single session.

Serial items were presented at a 2-sec. rate, paired associates at a 2:2-sec. rate, i.e., 2 sec. for stimulus and 2 sec. for stimulus and response. All learning and relearning took place by the anticipation method. The paired-associate lists were presented in three different orders to minimize serial associations.

For convenience in presenting the results, symbols for the four groups are as follows: M-S (massed, serial learning); D-S (distributed, serial learning); M-PA (massed, paired associates); D-PA (distributed, paired associates).

Results

Practice day.—The mean number of trials required by Group M–S to learn the practice list was 35.31 ± 2.10, and for Group D–S, 35.13 ± 2.10. *F* is less than one. In the paired-associate learning, Group M–PA took an average of 18.83 ± 2.40 trials to learn the practice list, and Group D–PA took 21.79 ± 2.40 trials. While these two means differ by nearly three trials, *F* is less than one. We may conclude that neither the two serial-learning groups nor the two paired-associate groups differed on initial ability.

Learning of experimental lists.—The mean number of trials to learn the experimental lists of all conditions is plotted in Fig. 1. The three degrees of learning are labeled low, moderate, and high. For the serial lists, Group D–S, with 30-sec. rest between each trial, is consistently superior to Group M–S. Analysis of variance has shown that the over-all difference between Groups M–S and D–S falls just short of the 5% level of confidence. In two

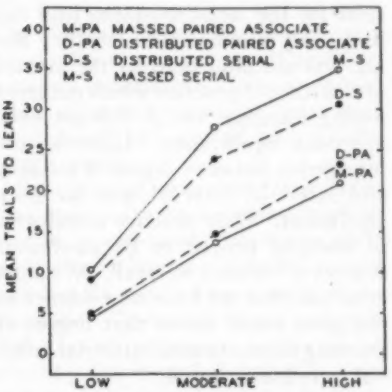


FIG. 1. Learning as a function of type of material, degree of learning, and intertrial interval

previous studies (4, 5) these same materials have been used with 30-sec. and 60-sec. intertrial intervals defining distributed practice. When evaluated against a massed condition, highly significant differences occurred in learning to one perfect trial. However, in these two previous experiments the same *S* served in both massed and distributed conditions. The statistical precision gained by such a design is greater than in the present experiments so that even though the quantitative differences between massed and distributed practice are of the same order in the present study as in the previous ones, they do not have as high statistical reliability. Nevertheless, as far as learning these serial lists of nonsense syllables is concerned, we may conclude that the present study shows what several previous studies have shown, namely, that distributed practice facilitates the learning of serial nonsense lists.

No appreciable difference between massed and distributed practice is evident in the learning of the paired-adjective lists. This finding also confirms two previous reports (3, 6).

Recall: serial lists.—The recall results for the serial nonsense lists can be quickly summarized verbally. Recall was not influenced by the massed or distributed practice which occurred during learning; the *F* did not even approach significance. Likewise, the interaction between degree of learning and intertrial interval was far from significant. Only practice and degree of learning proved to be significant sources of variance at recall. We must conclude that we have no evidence in the gross recall scores that degree of learning interacts with intertrial interval to influence recall.

Recall: paired-adjective lists.—For all degrees of learning, retention of paired-adjective lists was somewhat better following distributed learning than following massed learning, but the difference at each degree of learning was about the same. For the massed conditions the mean number of items recalled for low, moderate, and high degrees of learning was 1.50, 4.58, and 5.96 items, respectively. The corresponding values following distributed learning were 2.29, 5.25, and 6.63. Even so, the *F* for recall (3.45) falls short of the 5% level (4.05). Statistically speaking, the present recall results fall between two comparable studies (3, 6) in that the effect of distribution is less than in one and greater than in the other. The evidence therefore indicates that distributed practice of paired-adjective lists may be expected to produce a small but consistent superiority in recall over massed practice. Again, however, as in the case of serial lists, there is no evidence that degree of learning and intertrial interval interact at recall.

Relearning.—For both kinds of material the only variable which proved significant in relearning was degree of original learning. The greater the degree of original learning, the faster the relearning. Trends during relearning which occurred as a function of intertrial interval during original learning were comparable to those at recall, but statistically none of the differences was reliable.

It will be remembered that half the *S*s relearned by massed practice and half by distributed practice. Thus, of the groups which had original learning under massed practice, half had relearning by distributed and half by massed. The question is whether or not this change in context, i.e., change from massing to distribution or distribution to massing, influences relearning. The answer is clearly negative. Neither kind of material showed that this change was a significant source of variance; there were even no clear

trends which can be reported. In previous experiments (5, 7) serial nonsense lists originally learned by massed practice were more quickly relearned than those learned by distributed practice. This fact was very apparent with a 48-hr. retention interval and with lists of high intralist similarity. In these experiments relearning was always by massed practice. It thus might have been argued that the superiority of massing was due to the fact that the context of original learning and relearning was the same for material originally learned by massing, whereas this was not true for the lists originally learned by distribution. In view of the present findings that massed or distributed relearning seems to be of no consequence, it is quite unlikely that the previous results can be attributed to this change of context.

Other data bearing on the problem.—Our results have shown that degree of original learning does not interact with intertrial interval at recall. It would appear, therefore, that there is no confirming evidence for trends in previous findings that degree of learning is a relevant variable in assessing the influence of massed and distributed learning on retention. Among previous findings which suggested that it was a relevant variable were certain item analyses of recall (e.g., 5, Fig. 2) which showed that items given a large number of reinforcements during learning tended to be better recalled if this learning had been under massed rather than under distributed conditions. Conversely, items reinforced a small number of times were recalled better following distributed practice. Such evidence suggested that degree of learning was a relevant interacting variable. Having shown in the present studies that degree of learning is not a relevant variable when manipulated directly, the question still remains as to the interpretation of those differences which suggested the present experiments.

In the item analyses mentioned above, items which were reinforced many times are largely from *S*s who are slow learners. That is, the slow *S* gets a few items correct early in learning, and these are repeated correctly trial after trial while he tries to get other items correct. The fast-learning *S*s, on the other hand, acquire the entire list rapidly and most items are reinforced only a few times. These considerations, taken in conjunction with the results of item analyses as far as massed and distributed practice are concerned, suggest the possibility that whether massed or distributed learning is superior for recall depends upon the ability level of *S*. More specifically, it would appear that slow *S*s tend to recall better following massed practice than following distributed practice, and fast *S*s better following distributed than massed. In short, it might be that it is not degree of learning per se that is important, but the ability level of *S*s. To evaluate this possibility it is necessary to break down total groups into a minimum of two groups differing in speed of learning. We have done this for several previous sets of data. The division into two groups was always based on number of trials to learn all experimental lists combined when these experimental lists were learned both by massed and distributed practice.

Recall scores for fast and slow *S*s for the previous studies in this series show that in paired-associate learning, ability level and intertrial interval during learning do not interact appreciably at recall. There are suggestive trends that fast *S*s are more facilitated at recall by distributed practice than are slow *S*s, but until more definitive data are available this matter will not be pursued further. For serial learning, on the other hand, the trends

for both recall and relearning are very consistent; slow *Ss* show better retention following massed practice than following distributed, while fast *Ss* either show no difference or better recall following distribution. In the present serial-learning experiments, each group of 36 *Ss* was divided into two groups of 18 *Ss*, each based on speed of original learning. The recall of the fast *Ss* is superior following distributed practice for all three degrees of learning. For the slow *Ss*, on the other hand, recall is superior following massed practice for moderate and high degrees of learning. Taken separately, neither of these trends is significant statistically. However, because the trends were in the expected direction, we have made a thorough analysis of a previous set of serial experiments (5) in order to increase the number of cases in the slow and fast groups. These previous experiments involved the learning of serial nonsense lists of different levels of intralist similarity. A total of 108 *Ss* was used so that by ignoring intralist similarity we have 54 slow *Ss* and 54 fast *Ss*, each having a massed condition and two distributed conditions of 30 sec. and 60 sec. between each trial. A tabulation of the order of conditions for each group of 54 has shown that the two groups were almost perfectly balanced with regard to lists and order of conditions. Therefore, no bias should attach to either group because of differential practice effects or list differences.

The mean number of items recalled for these two groups for the three intertrial intervals is shown in Fig. 2. An analysis of variance has shown that interaction between ability level and intertrial interval is significant between the 5% and 1% levels of confidence (F is 4.17; F at 5% level is 3.04, and at 1% level, 4.71). We have also

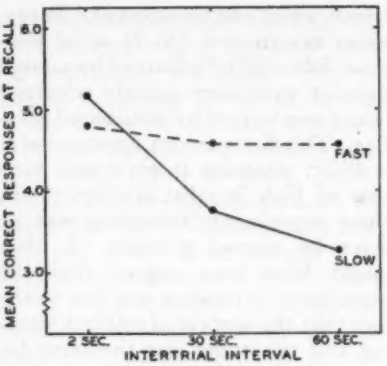


FIG. 2. Recall of serial nonsense syllables as a function of ability level and intertrial interval

made item analyses of recall in which number of reinforcements during learning is equated for massed and distributed conditions. Again, these data show that for slow *Ss* items learned under massed practice are better recalled than those learned under distributed practice. Fast *Ss* show little difference. These data, together with the trends exhibited in the present study and those of serial adjective experiments, lead us to conclude that recall following massed and distributed practice is differentially related to ability level.

DISCUSSION

As many studies have shown, degree of original learning and retention are directly related. But, the present results make it clear that whether retention is better following massed practice or following distributed practice in no way depends upon the degree of original learning per se. In neither paired-associate learning nor serial learning has degree of learning and intertrial interval shown an interaction at recall. The evidence which we first believed indicated that degree of learning and intertrial interval do interact at recall now appears to indicate that it is the ability level of *Ss* which is in-

volved, and not degree of learning. Following serial learning by massed practice retention will be better than following distributed practice for slow *Ss*. Fast *Ss* are more likely to show either better retention following distribution, or little difference in retention following massed and distributed practice. As indicated earlier, these relationships have appeared consistently in our studies involving serial learning but not in paired-associate learning. And, of course, we cannot specify in absolute terms what is meant by fast- and slow-learning *Ss*. All we have done to discover these differences in retention for fast and slow *Ss* is to divide our original group into two equal-sized groups based upon rate of learning.

The fact that ability level and intertrial interval interact in retention allows resolution of contradictions between Hovland's work (2) and our previous findings (5, 7). Hovland found that distributed practice was followed by much better retention than was massed practice. In our studies we have found either better retention following massed practice than following distribution, or no difference. In no case have we found that serial nonsense lists give better retention following distributed practice than following massed practice. The reason for this contradiction between Hovland's and our findings very likely lies in differences in ability level of the *Ss* used. Hovland's *Ss* as a group were more rapid learners than even our "fast" group shown in Fig. 2. On the average, our fast group required about 20 trials to learn a list of 13 syllables. Hovland's *Ss* required approximately 12 trials to learn a list of 12 syllables. These wide differences in learning ability may be attributed to the fact that Hovland's *Ss* were very well practiced, or that they had higher rote-learning ability, or both. Whatever the reason for these large differences in learning speed, it can be seen that on the basis of our evidence, Hovland's *Ss*, being very rapid learners, might be expected to show better retention following distributed practice than following massed practice.

The bases for the interaction in retention between intertrial interval and ability level will need much further study. We will need to know whether the differences in ability level must be "native" before the interaction occurs or whether initially slow *Ss*, after becoming very well practiced, will respond in the same fashion as initially fast *Ss*. Our data indicate that paired-associate lists do not show the interaction with ability level and intertrial interval that serial lists do. Yet, our data from paired-associate learning have shown trends which suggest that fast *Ss* will be more facilitated in retention following distributed practice than will slow *Ss*. This problem must be further explored by obtaining *Ss* with a greater range in ability level. It must also be noted that we have no way of telling from our data whether the differences in retention following massed and distributed learning for slow *Ss* result from a depression of recall following distribution or a facilitation following massing, or a combination of both. This, of course, makes the theoretical or explanatory problem a difficult one.

Superficially it might appear that a simple inhibition theory (e.g., reactive-type inhibition) would be appropriate to account for the differences in retention of fast and slow *Ss*. Slow *Ss*, taking many more trials to learn than fast *Ss*, would have generated much more inhibition under massed practice than under distributed practice. Following the 24-hr. rest, during which the inhibition dissipates, the items learned under massed practice would appear stronger than those learned under distributed practice. However, if this is the case we would expect that distributed practice would facilitate the learning of slow *Ss* much more than the fast *Ss*. Our analysis of the data on this matter shows this not to be true; fast *Ss* are facilitated as much by distributed practice as are slow *Ss*. Or, to say it another way, there is no interaction between intertrial interval and ability level during learning. It seems unlikely, therefore, that a simple inhibition theory will be appropriate.

Summary

Some evidence had suggested that whether or not distributed learning facilitates retention of

rote-learned materials depends in part upon the degree of original learning. Accordingly, a series of conditions was run to determine if intertrial interval and degree of original learning interact at recall. Two sets of materials were used, namely, serial nonsense lists and paired-adjective lists. One group of *Ss* learned serial lists under massed practice (2-sec. intertrial interval), and another group learned under distributed practice (30-sec. intertrial interval). With two additional groups the same conditions held for paired adjectives. Within each group three different conditions were used, these conditions differing only in degree of learning attained before the 24-hr. retention interval.

The results show:

1. Degree of original learning and intertrial interval during original learning do not interact during retention for either paired adjectives or serial syllables.

2. No significant differences were found in retention following massed and distributed learning, but all trends conform to previous findings in this series.

3. Change of context between original learning and relearning, change of context being defined as having massed practice during learning and distributed practice during relearning or vice versa, does not influence retention as measured by relearning.

4. Analyses of previous experiments as well as the present one indicated that degree of learning had appeared to be a relevant variable because differences in ability of *Ss is* a relevant variable differentially influencing recall following learning by massed and distributed practice. For serial learning slow *Ss* show better retention following massed learning than following distributed learning; fast *Ss* show no difference or better recall following distribution. For paired-associate learning no significant interaction occurs at recall between ability level and intertrial interval. These findings are used to account for existing discrepant results on the relation between distribution of practice and retention.

References

1. Greenberg, R., & Underwood, B. J. Retention as a function of stage of practice. *J. exp. Psychol.*, 1950, **40**, 452–457.
2. Hovland, C. I. Experimental studies in rote-learning theory: VI. Comparison of retention following learning to same criterion by massed and distributed practice. *J. exp. Psychol.*, 1940, **26**, 568–587.
3. Underwood, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42**, 153–161.
4. Underwood, B. J. Studies of distributed practice: VI. The influence of rest-interval activity in serial learning. *J. exp. Psychol.*, 1952, **43**, 329–340.
5. Underwood, B. J. Studies of distributed practice: VII. Learning and retention of serial nonsense lists as a function of intralist similarity. *J. exp. Psychol.*, 1952, **44**, 80–87.
6. Underwood, B. J. Studies of distributed practice: IX. Learning and retention of paired adjectives as a function of intralist similarity. *J. exp. Psychol.*, 1953, **45**, 143–149.
7. Underwood, B. J. Studies of distributed practice: XI. An attempt to resolve conflicting facts on retention of serial nonsense lists. *J. exp. Psychol.*, 1953, **45**, 355–359.

(Received August 10, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

THE PRINCIPLE OF SELECTIVE ASSOCIATION OF DRIVE STIMULI [1]

CHARLES B. WOODBURY AND DAVID H. WILDER

Oberlin College

The present experiment was designed to test, through two types of measures, Kendler's suggestion that "only those drive stimuli which are themselves reduced become connected to a rewarded response" (2; p. 217). The results partially confirm, in a complex maze situation, those of Levine (4).

METHOD

Experimental design.—Two groups of 20 naive, young male albino rats were given one trial per day on an elevated eight-unit multiple-U maze, with correct choices in the order RLLRLRRL. Both groups were alternately hungry (approx. 22 hr.) and hungry and thirsty (22 hr. each). The control group was rewarded with food on each trial; the experimental group, with food on hunger days and water on hunger-thirst days.[2] Kendler's principle would predict that the experimental group should learn more slowly than the controls, since the learning with one drive rewarded would not transfer to the alternate days when only the other drive was operative.

As each *S* reached the criterion of a total of four errors or less during three consecutive trials, he was started on a test series of one trial per day under thirst alone, with water reward. In this case, Kendler's principle would predict that the control group should show a greater decrement in performance than the experimental group; a nonselective or Hullian principle, on the other hand, would predict no such difference.

Apparatus.—The starting and transverse sections of the maze were 36 in. long, while the arms of each U were 24 in. long. The pathway was 1.25 in. wide. The maze ran diagonally across a room 18 ft. square, lighted by four overhead fixtures. In both training and test series, olfactory and other local cues were controlled by varying the order of running *S*s, by interchanging maze units, and reversing transverse sections daily.

RESULTS

Although running times were analyzed, they are not reported here as they parallel the error scores throughout. They may be found in the original study (5).

Table 1 provides a quantitative summary of the error data. Comparison

TABLE 1

MEAN SCORES DURING ORIGINAL LEARNING AND DURING SUBSEQUENT TEST UNDER THIRST ALONE

Score	Control Group* (N = 19)	Exp. Group (N = 20)	t	p
Trials to criterion†	19.3	20.3	.235	.10
Total errors†	79.1	81.1	.189	.10
Errors on all hunger days†	39.9	43.9	.781	.10
Errors on hunger Days 1–7	28.4	31.1	1.385	.10
Errors on last 3 days of training series	3.47	3.05	1.564	.10
Errors on last day of training series	1.0 ‡	.5 ‡	2.232	.02
Errors on first day of thirst test	4.4 ‡	1.9 ‡	4.378	.001
Error increase from last training to first thirst test day	3.4	1.4	4.000	.001

* Excluding results from one atypical animal that failed to reach criterion in 50 trials.
† Excluding criterion trials.
‡ Differences within these two pairs of means are both significant beyond the .001 level.

[1] Based on an M.A. thesis (5) by the junior author, directed by the senior author.

[2] To test the functional equivalence of the hunger and thirst components of the double drive state, a supplementary experiment was carried out in which two groups of ten rats each were maintained on the same dietary schedule as those in the main experiment. They were run only on double drive days, however, one group with food reward and one with water. The rate of learning of the two groups was almost identical: food group, 16 trials and 57.5 errors; water group, 17 trials and 53.5 errors. We conclude that the hunger and thirst components on the double drive days were functionally equal, under the deprivation and reward conditions used.

of the two groups with respect to number of trials to reach criterion, errors in learning, and errors on the hunger days alone (presumably a more sensitive test of Kendler's hypothesis) reveals no significant difference in the ease of learning the original maze habit under alternating drives.

Upon being tested under thirst alone, however, both groups showed significant increases in error scores, and the increase of the control group (never reinforced with water) was significantly greater than that of the experimental group.

Discussion

It is clear that the prediction based on the "selective principle" of association is not confirmed by these results, in the original learning phase of the experiment.

The relatively small number of errors made by the experimental group in the subsequent test under thirst alone, however, would seem to support the selective principle; it is, at any rate, consistent with it. This paradox raises the question of whether the difference between groups in the thirst test is actually incompatible with a nonselective principle such as Hull's.

It is our belief that this difference could be accounted for in terms of Hull's concept of the fractional anticipatory goal response (1, pp. 124 ff.) From this viewpoint, the anticipatory drinking responses, reinforced in the original training for the experimental group, could function effectively as cues during the thirst test; for the control group, however, such fractional anticipatory drinking responses would never have been reinforced during training and would be ineffective as cues later. The thirst test results may also be related to the Kendler and Levine finding (3) that thirst results in aversion to dry food. In view of these possibilities, it seems clear that these results cannot be advanced as strongly supporting the selective principle.

Summary

In order to test Kendler's "selective principle" of association between drive states and learned behavior, two groups of 20 male albino rats were trained on an eight-unit elevated multiple-U maze, one trial per day, under alternating hunger and hunger-plus-thirst drives. The control group was rewarded with food on all days; the experimental group, with food on hunger days and with water on hunger-plus-thirst days. After reaching criterion, each *S* was run under thirst drive alone, with water reward.

Contrary to prediction from the selective principle, the groups showed no difference in speed of learning the maze. In the thirst test, both groups showed a significant increase in errors over the last training trial. The control group, moreover, showed a significantly greater increase than did the experimental group.

It is concluded that these results offer no support for the hypothesis of selective association of drive states to responses.

References

1. Hull, C. L. *A behavior system.* New Haven: Yale Univ. Press, 1952.
2. Kendler, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
3. Kendler, H. H., & Levine, S. Studies of the effect of change of drive: I. From hunger to thirst in a T maze. *J. exp. Psychol.*, 1951, **41**, 429–436.
4. Levine, S. The role of irrelevant drive stimuli in learning. *J. exp. Psychol.*, 1953, **45**, 410–416.
5. Wilder, D. H. An experimental test of the selective principle of association of drive stimuli. Unpublished M.A. thesis, Oberlin Coll., 1951.

(Received August 3, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

TRANSMITTED INFORMATION AND THE LENGTH OF RATING SCALES

A. W. BENDIG

University of Pittsburgh

A previous paper by Bendig and Hughes (1) has suggested that information theory may be profitably applied to problems in constructing rating scales. The stimuli-rater-scale combination can be viewed as a communication channel with the stimuli providing the input information for the system, the rater acting as the transmission channel, and the scale operating as a decoder to yield output information. This analogy suggests many research problems concerning the effect of constraints acting on this channel at various points upon the transmission of information by the system. It is obvious that constraints applied at the input source, i.e., restricting the amount of information per stimulus, should reduce the average amount of information transmitted per response. However, constraints implicit in the rater, such as his channel width (the number of different types of symbols that can be transmitted), transmission speed (number of symbols per unit of time), and average amount of noise per symbol (rater error), will impose upper limits upon channel capacity, but the form of these functional relationships are, at this point, unknown. Constraints at the response (output) end of the channel, such as the number of possible alternative responses (length of scale) and the adequacy of scale anchoring in restricting total response information to stimulus-relevant information, may also influence channel capacity.

Bendig and Hughes (1) also presented data relating to the effect of scale constraints upon the transmission of information. Scales varying in the number of categories (3, 5, 7, 9, or 11) and in anchoring conditions (center category verbally anchored, both end categories anchored, or center and end categories anchored) were used by 225 raters to rate their knowledge of 12 foreign countries. The mean information transmitted per rating increased with greater anchoring of the scale. Transmitted information increased apparently linearly as a function of the number of scale categories, although a slight deceleration of the increase was noted between 9 and 11 categories.

This above-noted decelerated increasing relationship between transmitted information (I_t) and the number of scale categories (R) may be explained on the basis of a rectilinear relation between I_t and log R.[1] This appears plausible on rational grounds since log R, in the terminology of Garner and Hake (2), represents the maximum amount of information per symbol (Maximum Response Information or Maximum I_r) that a scale, used with a noiseless channel, is capable of transmitting. Theoretically it appears likely that the amount of information actually transmitted per symbol is a constant fraction of the maximum amount that could possibly be transmitted. If correct, this formulation may be somewhat analogous to the concept of the signal-noise ratio in communication research. However, the rectilinear relation of I_t and

[1] Logarithms referred to in this paper are to the base of two.

log R, inferred a posteriori from the previous study, needs to be checked in a new sample.

Experiment I

Method

Scales.—Numerical rating scales with 2, 3, 5, 7, and 9 categories were constructed and mimeographed along with rater instructions and the stimuli to be rated on single sheets. For the 2-category scale the instructions to the rater were those used by Wallen (8, p. 78). For the longer scales the instructions were modified to request *Ss* to rate each of the stimuli using the scale provided on the page. Three anchoring statements were used to define the center and both end categories of the scales of length 3, 5, 7, or 9. The anchoring statements were:

1. I like this food very much and eat it often.
2. I am somewhat neutral toward this food, neither liking nor disliking it much.
3. I dislike this food so much that I refuse to eat it.

The stimuli to be rated were the 20 foods used by Wallen (8, p. 80). These foods had originally been selected from a list of 142 foods (7) using the criteria that the foods were familiar to at least 90% of a group of 545 college *Ss* and disliked by not more than 10% of this group.

Subjects.—The *Ss* were 249 students in introductory and social psychology classes. The ratings of 13 *Ss* were excluded from the analysis because *S* used less than one half of the available categories in rating the foods. Thus, an *S*, using the 5-category scale, who used only ratings of one and five was not included.

Results

The information analysis followed the procedure (Method I) detailed by Garner and Hake (2), using the convenient table of $p \log p$ provided by Newman (6). This analysis was performed on each of the five lengths of rating scales and the data can be found in Table 1. As expected, transmitted information increased with longer scales from a value of .06 bits with the 2-category scale to .34 with the 9-category scale.

The original hypothesis of this investigation suggested that the linear correlation of the transmitted information values given in Table 1 with log R values would be extremely high and statistically significant. The correlation of I_t with log R was .950, this coefficient being significantly different from zero at the .05 level. The regression equation for predicting transmitted information from the logarithm of the number of scale categories was computed and found to be:

$$I_t = .114 \log R - .019 \quad r = .950$$

Discussion

The hypothesis suggested in the introduction to this paper, that the information transmitted by a rating scale is a rectilinear function of the maximum possible transmission, appears to be confirmed by the high correlation between I_t and log R. However, the relation of I_t and R between the limits of $R = 3$ and $R = 9$ is *essentially* a rectilinear relation, and it is only when shorter or longer scales are used that the decelerated na-

TABLE 1

Information Analysis of Rating Scales Varying in Number of Categories on the Scale

Number of Scale Categories	Number of Stimuli	Stimulus Information	Maximum Response Information	Obtained Response Information	Response Equivocation	Transmitted Information
2	20	4.32	1.00	0.47	0.41	.06
3	20	4.32	1.58	1.40	1.20	.20
5	20	4.32	2.32	2.12	1.85	.27
7	20	4.32	2.81	2.62	2.35	.27
9	20	4.32	3.17	2.98	2.64	.34

ture of the increase in I_t with greater R becomes evident.

As an interesting comparison, the correlation and regression equations were computed between the mean I_t and log R data reported by Bendig and Hughes (1). The ratings of 12 foreign countries by 225 college students using scales with 3, 5, 7, 9, and 11 categories gave the following correlation and equation:

$$I_t = .126 \log R + .143 \qquad r = .983$$

A comparison of these two equations suggests several interesting hypotheses. Both of the correlations are quite high (.950 and .983) and support the hypothesis that I_t is highly and consistently related to log R. The similarity in the slopes of the equations (.114 and .126) and the greater dissimilarity in equation constants (−.019 and .143) leads to the suggestion that for college *S*s (at least) the rate of gain of information with longer scales is relatively invariant, but that the absolute amount of information per scale length is related to some aspect of the rated stimuli. In an analogous sense, one minus the slope of the equation might be defined as rater "noise," i.e., the relative loss of information in the stimuli-rater-scale communication channel. Constraints on the "width" of this channel at the scale end, then, do not appear to greatly influence the relative magnitude of noise in the channel, but do influence the absolute amount of loss in transmission. In comparing short and long scales, more information is transmitted with longer scales, but the amount actually transmitted of the amount that theoretically could be transmitted by the scale remains constant.

One difference between the above two stimulus lists that could possibly account for the difference in equation constants is the amount of stimulus information (input) represented by each list. Stimulus information (I_s) is usually varied in three ways, i.e., by varying the number of possible stimuli presented to *S*, by varying the probability that any given stimulus will be presented, and by varying the sequential probability that one particular stimulus will be followed by some other stimulus. These three methods of varying I_s were used by Hyman (3) in his study of the relation of I_s and verbal reaction time. Essentially the same linear relationship between I_s and reaction time was found for each of the three methods used to vary I_s. However, a fourth method of varying I_s may be more applicable to the rating situation. As pointed out by Garner and Hake (2), each stimulus (S) represents a range of possible stimuli or events (E_1, E_2, E_3 . . . E_n) along an underlying event continuum. Stimulus information, then, is information about this event continuum. It can be argued that the width of the event continuum represented by the heterogeneity of the stimuli selected is also an important determinant of stimulus information. The researcher selecting stimuli to be rated acts as an encoding device, determining that a certain series of events shall be represented by a single message, the stimulus. If a heterogeneous set of stimuli is selected, then these stimuli represent a wide range of the event continuum, and each single stimulus (S) represents many possible events (E's). The problem for the rater in rating each stimulus is one of saying that this large number of events (represented by the stimulus) is the same as, or different from, the events represented by each category on the rating scale. Such a discriminative response on the part of a rater provides us with many comparisons between events and consequently yields much information about single event differences along this wide event continuum. However, for homogeneous stimuli each rater discrimination gives fewer event comparisons and provides less information about the total event continuum.

The above analysis suggests that if stimulus lists varying in the heterogeneity of the stimuli were rated using scales varying in length, I_t would be an increasing linear function of log R (or Max I_r), the slopes of these equations would fluctuate only because of sampling error, and the magnitude of the equation constants would be related to the stimuli heterogeneity of the lists. Corollaries of these hypotheses are

TABLE 2

INFORMATION MEASURES DERIVED FROM THE RATINGS OF STIMULUS LISTS

Stimulus List	Scale Length	Number of Raters	Maximum Response Information	Obtained Response Information	Response Equivocation	Transmitted Information
1	2	24	1.00	.70	.54	.16
	3	25	1.58	1.40	1.20	.20
	4	22	2.00	1.82	1.51	.31
	5	23	2.32	2.19	1.83	.36
2	2	20	1.00	.78	.71	.07
	3	25	1.58	1.39	1.17	.22
	4	21	2.00	1.81	1.56	.25
	5	21	2.32	2.15	1.87	.28
3	2	25	1.00	.62	.56	.06
	3	27	1.58	1.35	1.21	.14
	4	24	2.00	1.86	1.69	.17
	5	22	2.32	2.08	1.82	.26

that the absolute amount of I_t transmitted by the scales would increase with longer scales and also with increasing stimulus heterogeneity.

EXPERIMENT II

Method

Stimuli.—In the previous study 236 *S*s had rated the list of 20 foods as to preference value. From their mean ratings these foods were ranked in order from the most liked to the least liked food. Three sublists each containing 10 foods were then selected for the present study. List 1, the list containing the most heterogeneous food stimuli, was composed of the top five and bottom five foods from this ranking. List 2, of intermediate heterogeneity, was composed by selecting, in a double alternation pattern, 10 foods from this ranking. Foods ranked 1, 4, 5, 8, 9, 12, etc. were used for List 2. List 3, with the most homogeneous stimuli, contained the middle 10 foods: those ranked from 6 to 15. All three lists had a mean rank of 10.5 in the original ranking, but rank variances of 58.25, 34.25, and 8.25. The original list of 20 foods had, of course, a mean rank of 10.5 and a rank variance of 33.25.

Scales.—Four lengths of scale were used, containing 2, 3, 4, or 5 categories. The three descriptive statements used in the previous study were used to verbally anchor these scales. Scales with 3 or 5 categories had an anchor under each of the end categories and also under the center category. The 4-category scale also had an anchor under each end category, but the center statement was located midway between the two center categories. For the 2-category scale the center anchor was omitted, and the two end anchoring statements used with the other three scales were placed under the two categories. The lowest category on each scale was given a numerical weight of 1, the highest category numbered 2, 3, 4, or 5, with intermediate categories numbered accordingly.

Subjects.—The 12 combinations of three stimuli lists and four lengths of scale were mimeographed with instructions on single sheets and randomly distributed to 278 *S*s. The *S*s were students enrolled in daytime sections of introductory, social, applied, and educational psychology classes. The raters were told that *E* was investigating the adequacy of different rating scales in assessing the food preferences of college students.

Results

Information measures for each of the 12 subgroups (three stimulus lists by four scale lengths) were computed, and these measures are found in Table 2. Linear regression equations relating I_t to log R were derived separately for each of the stimulus lists and were found to be:

List 1 $I_t = .160 \log R - .0185$ $r = .974$
List 2 $I_t = .159 \log R - .0693$ $r = .962$
List 3 $I_t = .143 \log R - .0892$ $r = .980$

Our original hypotheses predicted that (*a*) there would be a positive increasing relationship between I_t and log R for each of the three lists; that

(*b*) when regression equations were fitted to the data from each list, the three equation constants would differ significantly, with the list containing the most heterogeneous stimuli, List 1, showing the largest constant; and that (*c*) the slopes of the three regression lines would not vary significantly. Visual inspection of the above equations confirms these predictions, but statistical tests of significance are necessary for adequate judgment. This immediately raises the as yet unsolved problem of the theoretical sampling distribution of information measures. In the absence of a derivation of the sampling distribution of I_t we might assume normality, but such an assumption appears highly hazardous. In the following statistical analysis we have applied nonparametric rank tests of significance that do not require assumptions about the form of the population distribution. However, it must be admitted that such tests of significance lack power and consequently increase the probability of Type II errors.

The first question to be answered was whether the slopes of the three equations were each significantly different from zero. The three product-moment correlations between I_t and log R were all significant at the .05 level, and this was confirmed by rank-difference correlations of 1.000 for each list between the same two variables. Since the equation slopes were all significant in a positive direction, then each list should show its smallest I_t value for the shortest scale length and its largest value with the longest scale. The four I_t values from each list were ranked in order of magnitude and the three ranks (one from each list) summed for each of the four scale lengths. Applying Kendall's coefficient of concordance (W) to the four sums of ranks (one for each length) the W value was 1.000 which is significant at the .001 level (5, p. 147). If either the slopes of the equations or the equation constants (or both) vary significantly among the three equations, we would expect the mean absolute values of I_t to vary significantly between the lists. To test this assumption the three I_t measures at each scale length (one from each list) were ranked and the sums of the four ranks (one from each scale length) were found for each list. The W value for the obtained sums of ranks was .812 which is significant at the .05 level (5, p. 146). This appears to establish that either the equation slopes or constants or both are significantly different from one equation to another. Finally, the significance of the variation in slopes was tested. If the slopes vary, the variability among the I_t measures from the three lists should be a function of the length of the scale. The four I_t measures from List 1 were paired with the four I_t measures from List 2 and the four differences obtained. These differences, taking regard to algebraic sign, were then ranked. Similar differences from the comparison of List 1 with List 3 and from comparing List 2 with List 3 were obtained and ranked. The sums of the three ranks (one from each comparison) were obtained for each of the four lengths of scale and W computed from these sums. W equaled .178 which is not significant at the .05 level (approximate probability of .80). It appears plausible to conclude that there is no relationship between the variability of I_t measures and scale length which logically leads to the conclusion of no significant variation in the slopes of the three equations. Therefore the significant variation in I_t measures among the lists must be attributed to significant variation in equation constants. Parametric and

more powerful tests of the differences between equation slopes and constants (**4**, pp. 88–91) were computed, and they gave results similar to those of the above rank tests.

Discussion

Our results appear to confirm the hypotheses given in the introduction to this paper. The three equations relating Max I_r and I_t describe a family of parallel straight lines with the same positive slope. Only the vertical displacement of the lines, represented by the equation constants, seems to vary as a function of stimulus heterogeneity. Since at present there is no method available for getting an absolute measure of stimulus heterogeneity, the form of the functional relationship between stimulus heterogeneity and the equation constant cannot be determined. In the absence of empirical data defining this relationship we might assume it to be rectilinear, but such a simplifying assumption needs experimental test.

One additional assumption has been used in this study, i.e., that I_t is a function of Max I_r and not of the obtained I_r. In all three studies obtained I_r has been very similar to, but less than Max I_r (or log R), since the raters tended to use each of the response categories equally often. Relating I_t to obtained I_r in these three experiments yields equations whose slopes are approximately the same as those reported, but whose constants are slightly smaller. This assumption implies that the possible information lost by the discrepancy between Max I_r and obtained I_r is "noise" or criterion-irrelevant information and that such a loss of possible information does not affect I_t or stimulus-relevant information. If correct, such an assumption suggests an important method of reducing relative "noise" (E_r) in the rating situation, i.e., deliberating varying scale anchors to reduce the obtained I_r. However, if the assumption is not valid, then such a reduction in obtained I_r would also reduce the obtained I_t. The assumption seems to be important enough both practically and theoretically to demand experimental test.

Summary

Two studies are reported applying information theory to rating scales. In the first study rating scales with 2, 3, 5, 7, and 9 categories were used by 236 *Ss* to rate preference values of 20 foods. The information transmitted by each scale length was computed. Transmitted information showed a significantly high correlation with the logarithm of the number of scale categories. It was concluded that the amount of information actually transmitted by a rating scale is a constant fraction of the maximum amount that could be transmitted under noiseless conditions. The second study checked this relationship when the input to the rater is stimulus lists varying in stimulus information. Three lists of ten foods were rated by 279 *Ss* using 2-, 3-, 4-, or 5-category rating scales. Foods for each list were selected so as to vary the heterogeneity of stimuli on each list. Information measures were computed and equations relating Transmitted Information (I_t) and Maximum Response Information (log R) found for each list. The obtained equations (*a*) showed a rectilinear positive relationship between I_t and log R, (*b*) did not vary significantly in their slopes or rate of increase, and (*c*) demonstrated a positive relation between the magnitude of equation constants and the stimulus heterogeneity of the lists.

References

1. Bendig, A. W., & Hughes, J. B. The effect of the amount of verbal anchoring and number of rating scale categories upon transmitted information. *J. exp. Psychol.*, 1953, **46**, 87–90.
2. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
3. Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45**, 188–196.
4. Johnson, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.
5. Kendall, M. G. *Rank correlation methods*. London: Griffin, 1948.
6. Newman, E. B. Computational methods useful in analyzing series of binary data. *Amer. J. Psychol.*, 1951, **64**, 252–262.
7. Wallen, R. Sex differences in food aversions. *J. appl. Psychol.*, 1943, **27**, 288–298.
8. Wallen, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.*, 1945, **40**, 77–81.

(Received July 9, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

TRACKING AND FREQUENCY OF TARGET INTERMITTENCE [1]

W. F. BATTIG, LEE W. GREGG,[2] E. H. NAGEL, ARNOLD M. SMALL, JR., AND W. J. BROGDEN

University of Wisconsin

Maximum proficiency in tracking will occur only when the operator positions the indicator so that it matches exactly the target position for every given instant of the course. This level of performance can be achieved only in theory. Target acquisition for an errorless run requires complete elimination of initial error by positioning movements that place the indicator on target and maintain it there. Errorless performance can be maintained only if rate of change in position of the target is predictable. With unpredictable target motion, error will occur. The operator can return to the on-target condition only after perceiving the magnitude and direction of error, and by adjusting his tracking movements. Thus, error is the rule. It also is the source of information by means of which the operator responds to reduce error. There is a continuous interaction between discrimination of error and adjustment of the tracking controls. There are time limitations, however, on this process. An error must reach a certain magnitude before it is perceived. This takes time, as do also the responses designed to correct the error. Maximum proficiency must therefore be defined as minimum error of tracking, and the role of error in its achievement must be recognized.

A variable that may contribute to faster learning and increased ultimate proficiency is intermittence in the presentation of the target. If the target is continuously viewed, and particularly if there is always an observed error, *S* will make continuous adjustive movements. He may increase the magnitude of his error by the continuous attempt to correct for error. It is also possible that this condition will result in a relatively high frequency of oscillation in the magnitude of tracking errors. If the target is present only intermittently, *S* can observe error only when the target is present and make adjustments only in terms of temporally limited observations. During the period when the target is absent, adjustive movements will not be made, and tracking should continue smoothly during these intervals at rates previously established. If this is so, there should be some optimal frequency of target intermittence for both accurate and smooth tracking. Tracking proficiency should decrease on either side of this frequency, becoming poorer both for increasing and decreasing frequencies.

The purpose of the present investigation is to determine the relation between proficiency in tracking and frequency of target intermittence. The experiment also provides a basis for evaluating the effect of amount of practice on this relation.

[1] This research was supported in part by the United States Air Force under Contract Number AF 33(038)-23294 monitored by the 6561st Research and Development Group (Perceptual and Motor Skills Research Laboratory), Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[2] Now at Carnegie Institute of Technology.

Apparatus and Procedure

The standard Pedestal Sight Manipulation Test (PSMT) has been described by Melton (5). This apparatus was modified to permit control of the frequency of intermittent target presentation by means of an electronic light flash generator. The circuits and operation of this instrument are described by Roush and Hamburger (6). A Sylvania type R1131C glow modulator tube replaced the standard 150-w. bulb in the projector unit of the PSMT. The scoring area at the standard calibration point (5) was set at 12 mils (approximately ¼ in.) for both azimuth and elevation. The range band at the same calibration point was set at 6 mils (approximately 5/16 in.). Standard Electric clocks were used to record the scoring times in .001 min. Additional measures were obtained from counters that recorded the number of times *S* moved into and out of the scoring area. Separate counts were obtained for azimuth and elevation, and for azimuth, elevation, and range.

The 16 *S*s were male undergraduate students at the University of Wisconsin. They were paid 75 cents per hour for participating in the experiment. Since the experiment was conducted under conditions of low illumination, *S*s were required to wear red adaptation goggles for at least 5 min. prior to entering the experimental room. Several additional minutes of adaptation occurred before tracking proficiency was measured. Each *S* received one trial of eight attacks for each of the eight frequencies of target intermittence in each of the four practice sessions.

At the first session, the instructions were read. They were not read again, but *S* was reminded of the nature of his task at each practice session. The instructions described the sight and the controls, and *E* demonstrated to *S* how to track and frame. Scoring was explained and *S* was informed that the target would go on and off intermittently. He was instructed to continue movement of the sight and range knob in anticipation of the course of the target. The kneeling platform was adjusted and questions were answered. The *S* removed his goggles and *E* left the experimental room to operate the control apparatus for the first trial of eight attacks.

A single 8 × 8 latin square was used for each of the practice sessions. The eight conditions of target intermittence were .4, .7, 1.0, 2.1, 4.1, 6.0, 8.2, and 16.1 cps. Limitations of the light flash generator made it impossible to provide the desired 50/50 light-dark ratio at all frequencies. The ratios were 20/80, 31/69, 47/53, 50/50, 49/51, 50/50, 50/50, and 41/59, respectively, for the different frequencies. Also, brightness of the target was not constant, but was a function of frequency (1).

The experiment actually made use of two copies of the PSMT and two *S*s were tested simultaneously. Since it was not possible to maintain the required calibration standards with the second apparatus, the results for the two devices differed and the data could not be pooled. Although the results obtained with the second apparatus are similar to those obtained with the other, they are not presented here, in order to conserve space and simplify the presentation. The requirement of testing two *S*s simultaneously and the failure of some *S*s to keep initial appointments for Session 1 resulted in an unequal number of *S*s for the eight sequences of target conditions. Satisfactory schedules, however, were maintained for Practice Sessions 2, 3, and 4. Only the results obtained in these sessions are presented.

Results

The measures of the amount of time each *S* simultaneously scored in azimuth and elevation (AE) and azimuth, elevation, and range (AER) were analyzed. Analysis of the counter measures of the number of times on and off target, and analysis of derived scores based on both time on target and counter measures, yielded essentially

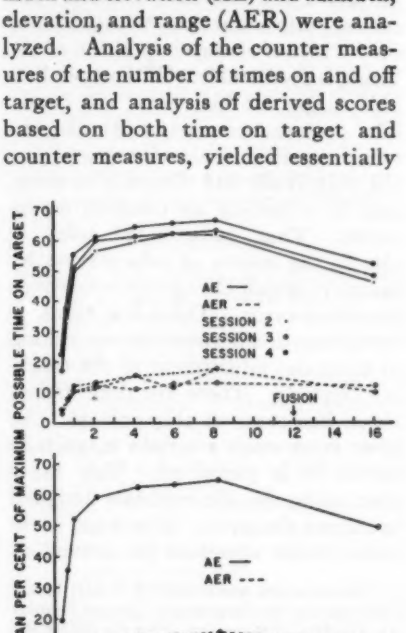


Fig. 1. Sighting proficiency in relation to frequency of intermittent presentation of the target and to practice

TABLE 1

SUMMARY OF ANALYSES OF VARIANCE OF AE AND AER SCORES

Source	*df*	AE			AER		
		Mean Square	V_2*	*F*	Mean Square	V_2*	*F*
1. Target conditions	7	3.408839	8	134.98**	0.201539	8	26.67**
2. Ordinal position	7	0.031031	8	1.23	0.003970	8	0.50
3. Sequence	7	0.720790	9	4.38**	0.139416	8	17.68**
4. Practice session	2	0.184999	8	7.33**	0.007580	8	0.96
5. PS × TC	14	0.003451	8	0.14	0.008265	8	1.05
6. PS × OP	14	0.004123	8	0.16	0.002293	8	0.29
7. PS × S	14	0.021496	10	1.68	0.007154	10	0.41
8. Square uniqueness	42	0.025255	11	3.43**	0.007886	11	1.68**
9. *Ss* within seq.	8	0.164510	10	12.85**	0.012274	10	0.43
10. PS × *Ss* with.	16	0.012803	11	1.74**	0.039958	11	8.50**
11. Pooled error	252	0.007362	—	—	0.004701	—	—

* Mean square used as denominator in *F* ratio.
** Significant at the .05 level of confidence.

the same results as did the time-on-target scores. For this reason, the results of neither of these latter analyses are presented. Figure 1 presents plots of the functions between tracking proficiency (in terms of mean AE and AER measures converted to percentage of maximum possible time on target) and target conditions. The upper half of the figure gives the curves for the individual practice sessions, and the lower half gives the curves based on means for all sessions. The shape of the functions is comparable for all conditions plotted. Tracking proficiency is poorest at the lowest frequency; it increases rapidly with increase in frequency to a maximum at 8.2 cps and then declines at 16.1 cps, the steady state target condition above fusion frequency. The function is steeper and shows greater differences between target conditions for AE than for AER. Level of proficiency for the more difficult AER performance is lower than for the AE. The AE curves for sessions show a practice effect, but those for AER show little or none.

The procedure for analysis of replicated latin-square designs has been described by Grant (4) and by Edwards (3). Each of the analyses is concerned with three repetitions of an 8 × 8 latin square with two replications. The three repetitions of the square represent the second, third, and fourth practice sessions. The raw scores are the clock scores in .001 min. for each *S* for each trial. Since two *Ss* completed each of the eight sequences, there are a total of 128 measures in each replication of the square or a total of 384 measures in each over-all analysis involving three repetitions of the latin square.

A summary of the results of analysis of variance of the AE and AER scores is given in Table 1. Grant (4) has pointed out that when there has been replication of a latin square, two estimates of error are available. One is the square uniqueness mean square, the usual error mean square for a single replication of the latin square. The other is the residual or the difference between the total sum of squares and the sum of the sums of squares for all other sources. The former term is subject to inflation or deflation by the unique pattern of confounding which has occurred in the interactions of the particular latin square. A test of this is available from the *F* ratio of these

two mean squares. Although the occurrence of significant inflation or deflation may be determined, it is not possible to discover which interaction or interactions are the source. If the square uniqueness mean square is inflated or deflated, the reverse influence will be present in one or more of the main effects. In such cases, therefore, the more conservative analysis will be that in which the larger mean square is used as the denominator of the *F* ratios.

Since the square uniqueness mean square is significantly greater than the pooled error mean square for each of the two analyses, it has been used as the denominator in testing the significance of the main effects. Target conditions (variation in frequency of target intermittence in cps) is significant at the 5% level of confidence for both AE and AER. Duncan's range test (2) was used to determine the reliability of the differences between the all session means for both AE and AER. Of the means for AE, those for .4 and .7 cps differ significantly from each other and from all other means. The means for 1.0 and 16.1 cps do not differ significantly from each other, but do differ significantly from all other means. The means for 2.1, 4.1, 6.0, and 8.2 cps do not differ significantly from each other. Of the means for AER, those for .4 and .7 cps do not differ significantly from each other, but do differ significantly from all other means. The mean for 8.2 cps differs significantly from the means for 1.0 and 16.1 cps. There are no other significant differences for AER.

In no case is the effect of ordinal position of target conditions significant. There is no reliable change in tracking between Trials 1 and 8 within the practice sessions.

The sequences mean square should be tested against the *Ss* within sequence mean square as well as against the error term, in order to determine whether a significant effect for sequences may be accounted for in terms of significant variation between *Ss*. This is so for the AER analysis, but not for the AE analysis. Examination of the row means for the AE measures shows no systematic pattern. The sequences for which the higher scores in sighting proficiency occur do not appear to differ in any systematic way from those for which lower scores occurred.

Session is significant only for the AE scores, indicating a reliable practice effect over Sessions 2, 3, and 4. Although there appear to be differential gains in practice as a function of frequency of intermittent presentation of the target, they are not reliable, since the interaction of sessions and conditions is not significant. The interactions of session and ordinal position and session and sequence are not significant for either of the analyses.

Discussion

The relation between frequency of intermittent target presentation and tracking proficiency is of particular interest since maximum scores were obtained, not with the steady state target (16.1 cps), but with an intermittent target at a frequency of 8.2 cps. These results suggest that maximum sighting proficiency may be achieved by intermittent viewing of the target. However, it is not possible to draw this conclusion even for the limited conditions of the present experiment, because of the confounding of light-dark ratio and target brightness with frequency. The light-dark ratio was 20/80, 31/69, 47/53, 50/50, 49/51, 50/50, 50/50, and 41/59, respectively, for target frequencies of .4, .7, 1.0, 2.1, 4.1, 6.0, 8.2, and 16.1 cps. There are significant departures from the desired 50/50 ratio, therefore, at target frequencies of .4, .7, and 16.1 cps. It is unlikely that there

was any effect for 16.1 cps, since this target condition was perceived as steady even though the light-dark ratio was 41/59. If it is assumed that tracking proficiency increases as the proportion of the light phase is increased up to 50%, the increments in the obtained curves from .4 to .7 to 1.0 cps may be a function of the combined effects of increase in frequency of intermittent presentation of the target and increase of light phase, or a function of increase in light phase alone. The significant increment in the AE curve from 1.0 to all frequencies between 2.1 and 8.2 cps, and the unreliable but progressive increments from 2.1 to 4.1, from 4.1 to 6.0, and from 6.0 to 8.2 cps, however, can be attributed only to change in frequency of the target.

It was not possible to measure absolute brightness of the target for the different frequencies. Relative measures, however, were obtained in terms of filter density necessary to make an adjacent spot of light equal in brightness to the target. Target brightness decreased linearly with increase in frequency. The decrease between .4 and 16.1 cps is on the order of 1 log unit. If brightness affects tracking at all, the expected relation would be positive. Assuming such a relationship, its effect on the obtained curves would be deflation of the increments in tracking with frequency, and inflation of the decrement between 8.2 and 16.1 cps. In terms of the form of the obtained function, only the decrement from 8.2 to 16.1 cps is critical. It might be partially or wholly a function of the decrease in brightness.

Thus, the curves obtained between frequency of intermittent presentation of the target and proficiency of tracking may or may not represent simply the relation between these two variables. Without knowledge of any relation between light-dark ratio and tracking, or between target brightness and tracking, however, the most reasonable interpretation of the curves is that they represent approximately a functional relation between proficiency in tracking and frequency of intermittent presentation of the target.

Only further research on tracking in which light-dark ratio and brightness are adequately treated as parameters will solve the problems inherent in the present exploratory study. Instrumentation for such research will be difficult to achieve. Further development of the light-flash generator will be necessary to achieve a wider range of light-dark ratios, including 50/50, at frequencies below 1 cps. Only glow modulator tubes may be used with such generators, because they are the only type of lamp now available that responds immediately and faithfully to the output of the generator. The maximum brightness of such tubes is low, and because of this, it was necessary in the present experiment to work at relatively low frequencies and to dark adapt the *S*s. Future studies, therefore, will have to be made at low frequency levels and low levels of brightness, unless lamps of much greater intensity are developed. The low brightness level complicates, but does not make impossible, the equation of target brightness. Bartley (1) has shown that the subjective brightness of a transmitted flickering light increases to a maximum value almost twice that of the subjective brightness of a steady light of the same intensity. At critical flicker fusion, the subjective brightness of a 50/50 light-dark ratio, is only half that of a steady light of the same intensity. A determination of the Bartley function for the conditions of further study of tracking and frequency of intermittent target presentation will make it possible to maintain a constant level of brightness of the target.

Summary

Measures of proficiency of tracking on the Pedestal Sight Manipulation Test were obtained at frequencies of intermittent target presentation of .4, .7, 1.0, 2.1, 4.1, 6.0, 8.2, and 16.1 cps. Since fusion frequency was at approximately 12 cps, the target at 16.1 cps was perceived as steady. The 16 *S*s of the experiment, all male undergraduate college students, received one trial of eight attacks for each frequency in each of four experimental sessions.

Separate analyses of variance were completed for the AE and AER measures. A significant differential effect for frequency of intermittent

target presentation was obtained. The relation is that of a bowed curve concave downward. As frequency of target intermittence increases, there is an initial rapid rise in tracking proficiency to 1.0 cps, a more gradual increase to a maximum at 8.2 cps, and a decline at 16.1 cps. A significant practice effect was obtained only for the AE measures.

Since a light-dark ratio of 50/50 could not be obtained at .4, .7, and 16.1 cps, and since target brightness declined consistently as frequency increased, the obtained curves cannot be said to be a simple function of frequency of intermittent presentation of the target. The gains in tracking proficiency between 1.0 and 8.2 cps, however, are not differentially affected by the constant light-dark ratio of 50/50, and the only effect of target brightness could be one of decreasing the magnitude of the increments. There is, therefore, evidence for a relation of tracking and frequency of target presentation for which neither light-dark ratio nor brightness may be responsible. The obtained curves, however, will require confirmation under conditions that permit adequate control of light-dark ratio and target brightness as parameters.

References

1. Bartley, S. H. Subjective brightness in relation to flash rate and light-dark ratio. *J. exp. Psychol.*, 1938, **23**, 313–319.
2. Duncan, D. B. A significance test for differences between ranked treatments in an analysis of variance. *Virginia J. Sci.*, 1951, **2** (new series), 171–189.
3. Edwards, A. G. Homogeneity of variance and the latin-square design. *Psychol. Bull.*, 1950, **47**, 118–129.
4. Grant, D. A. The latin-square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.
5. Melton, A. W. (Ed.) *Apparatus tests.* Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4.)
6. Roush, R. G., & Hamburger, F., Jr. A light flash generator. *Electronics*, 1948, **21**, 100–102.

(Received July 17, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

RESPONSE TIME PATTERNS ASSOCIATED WITH VARIOUS DISPLAY–CONTROL RELATIONSHIPS [1]

W. D. GARVEY AND W. B. KNOWLES [2]

Naval Research Laboratory

In a previous experiment (2) in which several display-control systems combined from the *same* group of displays and controls were evaluated, it was determined that some combinations were more efficient than others. These results were discussed within the information theory framework, and were interpreted to mean that *S*s were able to utilize a simpler coding procedure to transmit information from display to control in operating the more efficient systems. Using a group of similar systems, the present study was conducted to determine: (*a*) if the relative efficiencies of the systems would remain the same throughout a period of practice, and (*b*) if a secondary task would have the same interfering effect on the operation of all systems. Furthermore, it was anticipated that (*a*) differential progress in learning to operate the different systems, (*b*) differential interfering effect by the secondary task on the operation of the different systems, and (*c*) an analysis of the component S–R elements would give some notion of the nature of the different coding procedures used to operate the various systems.

The over-all plan of the experiment was as follows: Two displays and two controls were combined in different ways to create six systems. These six systems were further classified into three pairs, so that the informational characteristics (6.64 bits per stimulus) and the motor difficulty were equivalent from *pair* to *pair*. However, by manipulating the degree of isomorphic relationship between stimulus set on the display and response set on the control, an attempt was made to vary from pair to pair the directness with which the stimuli denoted the responses. The three pairs are shown as Cond. I, II, and III in Fig. 1. The *S*s were given several weeks practice on each of the systems, after which a secondary task was introduced.

METHOD

Apparatus.—Detailed descriptions of the general apparatus have been reported (2). Two displays were used: (*a*) a 10 × 10 matrix of 100 lights, and (*b*) two columns of ten lights (see Fig. 1). Each display was capable of presenting 100 stimulus alternatives. A stimulus light on the matrix display could be identified in terms of a lettered row and a numbered column; e.g., the upper left corner stimulus light was coded "A–1." On the column-column display, two lights constituted a stimulus; lights in the left-hand column were lettered and lights in the right-hand column were numbered. Thus, each pair of lights could be identified by a letter-number code; e.g., the top two lights constituted stimulus "A–1." The physical make-up of the controls was equivalent to the displays: (*a*) a 10 × 10 matrix of push buttons, and (*b*) two columns of ten push buttons. As with the display lights, the response buttons could be coded in terms of letters and numbers. The *S*'s task was to press the control buttons which were identical in the letter-number code to the stimulus lights. The resulting systems are shown in Fig. 1 where Cond. I refers to the systems which had display and control combined into one panel in such a way that the light stimuli were directly above the response buttons. The S–R relationship was considered the least complex for these systems, the next least complex for Systems C and D, and the most complex for Systems E and F.

[1] The opinions or assertions contained herein are the private ones of the writers and are not to be construed as official or reflecting the views of the Navy Department or the naval service at large.

[2] The writers wish to acknowledge the valuable assistance of Mr. E. P. Newlin in collecting and analyzing the data of this experiment.

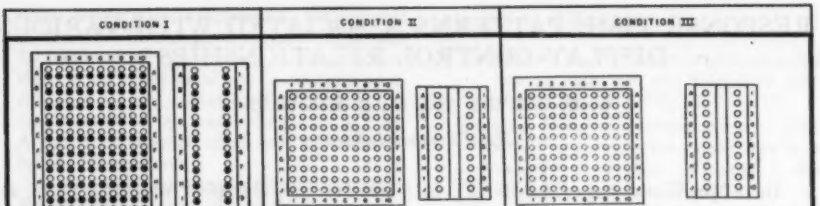


FIG. 1. Displays and controls. The hollow circles represent the display lights and the filled circles represent the control buttons.

The display-control systems were located in a sound-treated room remote from *E* and the main apparatus. The *S*s were seated at a small table on which the control panels were placed. The table was adjusted to a comfortable level and tilted about 35° towards *S*. Displays were centered vertically above the controls on a small shelf at approximately eye level and about 30 in. from *S*. In the case of Systems A and B the combined display and control unit was placed on this table. The walls and ceiling of the room were white; fluorescent ceiling lights provided a moderately bright general illumination.

Random sequences of stimuli were presented by means of a specially constructed, punched-tape, programming device. This apparatus also accumulated response times throughout the experimental trial of 100 stimuli. Simultaneous polygraph records were taken so that response times and errors to specific stimuli could be measured. The stimulus presentation was self-paced.

Procedure.—Six U. S. Navy enlisted men stationed at the laboratory were given one trial of 100 stimuli on each of the six display-control systems each day for 19 days. Six sequences of 100 stimuli, randomized with the restrictions that no two successive stimuli would have either the same number or letter code, were used. The experimental conditions and stimulus sequences were presented according to a number of counterbalanced orders which were changed from day to day. The efficiency of a given system was measured in terms of mean response time per stimulus. Throughout the experiment *S*s were reminded before each trial how the system was operated and that they should work as fast as possible *without* making errors. Error scores were recorded as a check that *S*s were not achieving speed at the expense of accuracy.

On Days 16 and 17 the primary experimental conditions were presented in the usual manner. In addition, *S*s were required to perform a second task while simultaneously operating the display-control systems. This task was to count and report aloud the number of clicks presented at intervals over a loudspeaker. The *S*s were instructed to concentrate on performing the secondary task perfectly even at the expense of the primary task. It was assumed under these conditions that the performance of the secondary task should increase the performance time of each of the six primary tasks by an equivalent amount, unless there were S–R elements of the secondary task that interfered with the S–R elements of some primary tasks more than with the others.

The secondary click stimuli were presented in groups of 4, 5, 6, or 7 clicks. Twenty such groups, five of each number, were randomly distributed throughout each trial by presenting a group every 4, 5, or 6 primary responses. The number of clicks in the groups and the intervals between the groups were appropriately randomized so that the same number of clicks was not presented successively and so that no *S* had the same series of secondary stimuli more than once. On the secondary task days, in addition to measuring the primary task response time, a record was kept of errors made on the secondary task.

Initially, it was planned to discontinue the experiment at the end of Day 18. However, since *S*s' tour of duty was extended and it was desirable to obtain more data for an analysis of the component S–R latencies associated with performance on Systems C, D, E, and F, the experi-

ment was continued. However, Systems A and B, for which sufficient data had been collected, were discontinued.

Results[3]

Relative efficiencies of systems.—The major results are summarized in the curves in Fig. 2. The mean response times per stimulus were obtained for each system. For the first 18 trials the systems ranked consistently from fastest to slowest mean response time in the following order: A, B, C, D, F, and E. For Trials 19 through 24 the systems ranked: C, D, F, and E. These differences between means were statistically significant beyond the 1% level when tested by Friedman's (1) ranking technique.

Errors were made to less than 3% of the stimuli. A statistical analysis using Friedman's technique (1) indicated that the difference between errors made on the various systems was not significant.

Effect of practice.—The curves in Fig. 2 illustrate the effect of practice on the performance. The trend of the mean response time as a function of practice (excluding Trials 16 and 17) was tested with Lindquist's Case 2 (3, p. 67). The trends of the means for Systems C, D, E, and F were found to be significant ($p<.01$); the trends of the means for Systems A and B were found not to be significant.

Using Lindquist's Case 12 (3, p. 78), it was found that the trends obtained for Systems E and F each differed significantly ($p<.01$) from those of Systems A, B, C, and D. The trends for Systems C and D differed significantly ($p<.05$) from those of A and B. However, tests of difference between the trends of Systems A and B, C and D, or E and F did not prove significant.

Effect of secondary task.—The general effect of the secondary task on the efficiency of the various systems is shown by the increased response times for Trials 16 and 17 in Fig. 2. To determine if the secondary task increased the mean response times significantly, matched t tests were applied to the difference between the means of Trials 15 and 16 and between the means of Trials 15 and 17. On Day 16, Systems A and B showed no significant increase; the increase was significant for Systems C and D ($p<.05$) as well as for Systems E and F ($p<.01$). On Trial 17 the increase over Trial 15 for System F only was significant ($p<.05$).

To test if the secondary task differentially increased the response times for the three conditions of display-control relationship, the differences in response times between Trials 15 and 16 were obtained for each of the sys-

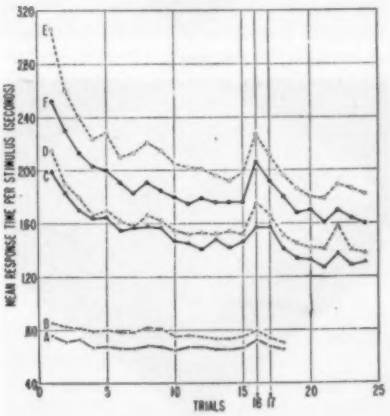


Fig. 2. Mean response times per stimulus (seconds) for the display-control systems on the 24 practice trials. One trial consisted of each of the 100 possible stimulus alternatives.

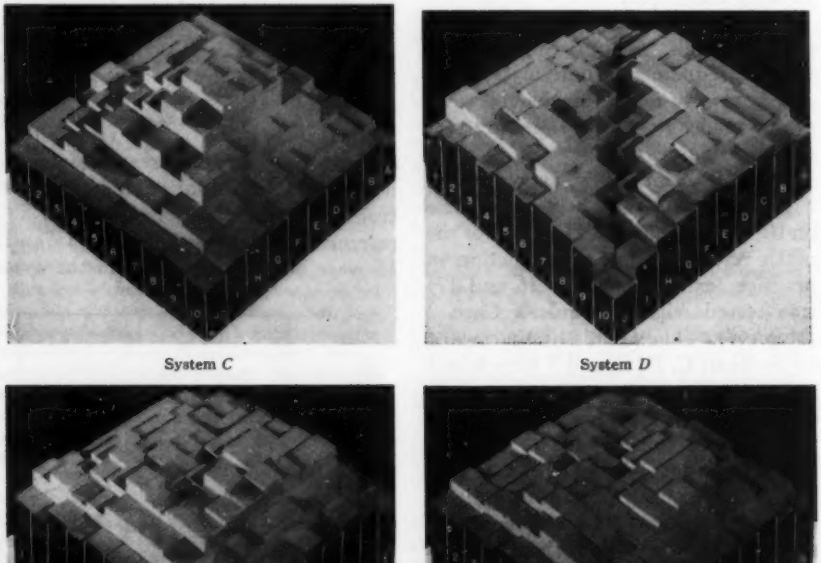
[3] A 19-page table giving critical summaries of data as well as the results of the analyses of variance has been deposited with the American Documentation Institute. Order Document No. 4200 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C., remitting in advance $2.50 for 35-mm. microfilm or $1.75 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

tems. The t tests on the pooled data indicated that the increase in response times for the systems in Cond. II differed significantly ($p < .05$) from those of Cond. I; differences in the increased response times between Cond. III and Cond. II were also significant ($p < .01$).

Statistical analysis (1) of the secondary task errors was made to determine if there was a tendency for more of these errors to be made while operating one primary system than another; no significant differences ($p > .25$) were found for either Trial 16 or 17.

Analysis of component S–R elements.—From the polygraph records, direct measures of response times to specific stimuli could be obtained. The records for Trials 1–4, 9–12, and 18–21 were analyzed, and a mean response time based on this sample of 72 readings was obtained for each S–R element.

Preliminary analysis of Systems A and B indicated that the response time latencies were about the same for all stimuli, and no reliable differences could be established. Therefore, a more extensive analysis of these two systems was not continued. However, definite patterns were apparent for Systems C, D, E, and F. The graphic models, which illustrate the



System *C* System *D*

System *E* System *F*

FIG. 3.· Graphic models of S–R latencies

different patterns, may be seen in Fig. 3. Since each stimulus could be identified in terms of a letter-number code, the models have been constructed in matrix form where the mean latency of an S–R element is identified in terms of the letter-number code of the stimulus. The heights of the bars represent the mean response times to the various stimulus elements.

It appears from these models that different patternings of response times are obtained for the various display-control combinations. To test the presence of a pattern, each response-time matrix shown in Fig. 3 was tested for a significant row-by-column interaction, using Wilcoxon's (**4**) ranking procedure and considering the mean response times for Trials 1–4, 9–12, and 18–21 as three replications of the matrices. The *z*-score equivalents for the χ_r^2 values were 4.43 or greater, indicating that the patterns were extremely reliable even when tested by this rather crude method. To test for the difference between any two patterns a matrix of the cell differences was constructed and the row-by-column interaction of this difference matrix was examined by the same test. The *z*-score equivalents for the χ_r^2 values were 4.84 or greater for all comparisons, except that of Systems E and F. Here the *z*-score equivalent was .39 ($p > .70$).

The salient features of these models appear to be the following. The model for System C shows a steep increase from the periphery to the center of the bar matrix. A similar peripheral-central gradient is apparent for Systems E and F, but the gradient is not so steep. The peripheral response times for System C are more of the order of the mean of response times for Systems A or B; whereas the central response times are more of the order of the mean of response times for Systems E or F. In addition, it may be noted that the peaks of the distributions appear to be slightly below and to the right of center.

The model of the response times for System D does not reveal the peripheral-central gradients found for the other systems. Here, the shortest response times are found along the major diagonal of the model. Within any particular row or column it can be observed that there is an inflected gradient in either direction from these diagonal points, increasing at first and then diminishing as the edges are approached to reach secondary minimum at the edges. This definitive patterning is not observed for Systems E or F even though they have one component (either the display or control) common to System D.

Discussion

Relative efficiencies of systems.—Since the same displays and controls were used in each of the three conditions of display-control relationship, it may be assumed in making comparisons between these conditions that the same amount of information was presented per stimulus and that the controls could be operated with equivalent motor difficulty. Therefore, the significant differences in performance time between conditions must be attributed to the differences in the complexity of the discriminal process. That is to say, some combinations permitted a more direct translation from stimulus to response, and manipulating the relationship between displays and controls has introduced "method of coding" as the variable underlying complexity of the discriminal process.

Effect of practice.—The amount of practice given *S*s did not change the relative efficiencies of the various systems. However, progress in learning the perceptual-motor skills involved in the operation of these display-control systems differed even though each of the tasks involved the same number of discriminated reac-

tions to the same number of stimuli. The data show that the fastest acquisition rates and smallest response time decrements were found for Systems A and B. It is therefore concluded that the generalization tendency was lowest among the learned reactions of these tasks. In other words, the stimuli in these systems denoted most directly the responses to be made, and *S*s were able to use the simplest method of coding the information.

The trend analysis indicated that the decrements in response time with practice for Systems C and D were greater than for Systems A and B. This is taken to indicate a greater degree of generalization in Systems C and D, and that the stimuli in these systems did not denote the responses as directly as in Systems A and B. This is interpreted to mean that cues in addition to or different from those of Systems A and B were required in order to discriminate the proper response. Thus, *S*s were required to use a more complex method of coding information to operate Systems C and D than to operate Systems A and B.

The performance times for Systems E and F showed the greatest decrements and slowest acquisition rates, indicating that the relationship between stimuli and responses in these systems was the least direct. This is interpreted to indicate that, in general, the method of coding used to operate these systems was the most complex.

Secondary task.—In an earlier experiment (2) *S*s reported that they had to make greater use of the letter-number indices (located on the displays and controls) to operate the less efficient systems (e.g., System E) than to operate the better systems (e.g., System C). This led to the speculation that *S*s were required to use more analytical coding procedures to operate Systems E and F than to operate the other systems. On the assumption that a secondary task which was analytical in nature would have the greatest effect on the operation of those systems which required analytical coding procedures, the click-counting task was selected as a secondary task. That the secondary task did not affect the performance on Systems A and B is interpreted to mean that the method of coding used to operate these systems was, in general, of a nonanalytical nature. The secondary task had a greater interfering effect on performance on Systems C and D than it did on Systems A and B; but it had a lesser effect on performance on Systems C and D than on Systems E and F. This is taken to indicate that the coding employed in the operation of Systems C and D was in general more analytical than that of Systems A and B, but less analytical than that employed in the operation of Systems E and F.

Analysis of S–R elements.—To arrive at a tentative explanation of the differences in efficiency between the systems in Cond. II and III, an attempt was made to determine what relationships the S–R latency patterns bear to the spatial configurations made possible by the specific combinations of displays and controls used in these systems.

The model for System C shows that the response times to stimuli located around the edges of the display were considerably shorter than for those located in the center. It seems reasonable to suppose that the edges and corners of the matrices provide spatial anchoring points for identifying these peripheral stimuli. It also seems reasonable to suppose that the farther a stimulus is away from these anchoring points, the greater will be the internal interference among the elements, and the longer will be the response latency. It would also seem possible that in the center of the display, where the spatial references are not so readily available, *S*s were required to utilize cues in addition to the spatial anchor points to make the proper response.

The patterning of the S–R latencies for System D is more complex, but it can be interpreted in much the same manner as that for System C. When it is remembered that in System D *two* lights constituted a stimulus and the response was made on *two* buttons, the pattern might be interpreted to indicate that three spatial anchoring points were used by *S*s. When either of the stimulus components (i.e., letter light or number light) was

near either the top or bottom of the columns, the response time to the stimulus was shorter. The response times were also shorter when the components were adjacent to each other. In each instance where one component of the stimulus could be anchored to one of these spatial points, it would seem possible for *S*s to utilize these spatial cues to identify the proper response.

Even though Systems E and F contain, either as display or control, components of both Systems C and D, the patterns for Systems E and F were distinctly different from those of C and D. However, response times obtained for Systems E and F were, in general, of the same magnitude as the most central ones in System C. It may be inferred from this that the coding employed in operating Systems E and F was, in general, of the same order of complexity as the most complex coding employed in operating System C.

To determine if the secondary task had a greater interfering effect on S–R elements considered not to have readily available spatial anchor points than on those considered to have such anchor points, the following analysis was made. An arbitrary classification of the S–R elements into "spatial" and "nonspatial" was made on the basis of the response-time patterns. In the model for System C, the outer three rows and columns of bars were termed "spatial"; the central 16 bars, "nonspatial." This particular breakdown is called the "matrix" classification. For System D the bars on the major diagonal and along the edges were termed "spatial"; this is called the "column-column" classification. The primary task elements were further categorized into "distracted" and "nondistracted," depending upon whether or not secondary task stimuli (on Trial 16) coincided with them.

Two populations of response times, one for spatial elements and another for nonspatial elements, were established for each system by pooling all the response times of all the subjects on Trial 16. A χ^2 test was then made to determine if the proportion of distracted, nonspatial elements above the median of total nonspatial population was greater than the proportion of distracted, spatial elements above the median of the total spatial distribution. These tests showed that relative shift of the nonspatial elements under distraction was significant in the cases where System C was analyzed according to the matrix classification and where System D was analyzed according to the column-column classification. However, no differential effect of distraction was found when Systems E and F were analyzed in terms of either classification. In short, all elements in both Systems E and F were equally and greatly disrupted; all the elements of Systems C and D were disrupted to some extent, but those which were classified as nonspatial were disrupted to a greater extent.

Complexity of coding.—On the whole the results of this experiment point out that the kinds of discriminations, or the methods of coding, that an operator utilizes in any display-control system are a function of the relationship between the display and control, and may be extremely important in limiting the efficiency of the system. In developing this argument the tasks and task elements have in various places been referred to as spatial, nonspatial, analytical, nonanalytical, etc. The experiment certainly has not demonstrated these terms to be exactly descriptive of the dimensions underlying complexity of coding. The results do, however, provide some clues which permit the speculation that additional research may help to tie down complexity of coding to perceptual-analytical, spatial-verbal, or iconic-symbolic dimensions.

Summary

Two displays and two controls were combined in different ways to create six display-control systems. These six systems were then classified into three pairs, so that the informational characteristics and motor difficulty were equivalent from pair to pair. The experiment was designed to determine if: (*a*) the various systems would maintain their same relative efficiencies throughout a period of practice, and (*b*) a secondary task would have the same interfering effect on the operation of all systems. At the same time it was anticipated that a detailed analysis of the

individual response-time data would provide some insight as to the nature of the coding used by *Ss* to transmit information from display to control.

The results were as follows:

1. Efficiency of performance (measured in terms of mean response time per stimulus) on the various systems differed significantly.

2. The relative efficiency of the systems did not change throughout practice.

3. The progress of learning the perceptual-motor skills involved in operating the various systems differed significantly between the three pairs.

4. A secondary task had no significant effect on the performance of the most efficient systems; it had the greatest interfering effect on the performance of the least efficient systems.

5. Different patternings of S–R latencies were obtained from performance on the various systems.

References

1. Friedman, M. Use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Amer. statist. Ass.*, 1937, **32**, 675–701.
2. Knowles, W. B., Garvey, W. D., & Newlin, E. P. The effect of "speed" and "load" on display-control relationships. *J. exp. Psychol.*, 1953, **46**, 65–75.
3. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1946, **12**, 65–78.
4. Wilcoxon, F. *Some rapid approximate statistical procedures.* Stamford, Conn.: American Cyanamid Co., 1949.

(Received July 23, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

SHAPE DISCRIMINATION AS A FUNCTION OF THE ANGULAR ORIENTATION OF THE STIMULI [1]

MALCOLM D. ARNOULT

Skill Components Research Laboratory, AF Personnel and Training Research Center

A systematic study of the task of shape perception involves investigating those variables which contribute to the perception of a shape in any given situation. For any given shape, or within any class of shapes, it is possible to manipulate size, brightness, color, orientation, familiarity, and perhaps many other stimulus attributes. It is also possible to make a rough classification of the possible perceptual tasks into discrimination, recognition, and identification, depending upon whether *S* is expected to (*a*) detect a difference between two shapes, (*b*) point out a familiar shape, or (*c*) give the proper name to a shape. The most difficult problem in shape-perception research is that of specifying the attribute of "shape" itself, and this problem will be discussed presently.

This report deals with the effect on simple shape discrimination of varying the angular positions of the stimuli. The stimuli were two-dimensional nonsense shapes which could be rotated in their own plane to any of eight different positions ("right-side-up," "upside-down," and six intermediary positions). In other words, they were rotated about their centers on an axis parallel to *S*'s line of sight, and perpendicular to the plane in which they were projected. The various combinations of positions of the two shapes in each pair resulted in 13 different amounts of angular separation. The accuracy and latency of shape discrimination were studied as a function of this angular separation. The experiment was similar, in general, to one reported in 1935 by Gibson and Robinson, except that they used familiar shapes and fewer angular positions (3).

In varying the orientation of nonsense shapes it is necessary to specify (*a*) the orientation of any one shape relative to some zero point, and (*b*) the conditions under which two or more shapes are considered to have the same orientation. As long as the shapes used can be given a determinate orientation (e.g., the orientations of circles, squares, and triangles would be wholly or partially indeterminate), the first requirement can be met by selecting an arbitrary zero point. The "normal" (or 0°) position in this experiment was so chosen. The second requirement was met in this experiment as follows: two shapes were judged to have the same orientation if superposition of them would result in a maximum of overlap of their areas. While this definition appeared to be adequate here, it would not necessarily be appropriate for a different group of shapes.

The selection of the shapes to be used as stimuli in an experiment of this sort poses what is essentially a sampling problem. Ideally, the results should be generalizable, not only to a population of observers, but also to a population of shapes (2). Presumably, maximum generalization of the latter type could be achieved by using a sample of shapes which was representative of the population of all possible shapes. The selection of such a sample is unfeasible at the present time because no method exists for measuring the attribute of "shape" in the sense that one can measure such attributes as weight, size, brightness, etc. It is not possible either to define the

[1] The data reported in this study were collected as a part of the United States Air Force Human Resources Research and Development Program. The opinions or conclusions contained in this report are those of the author. They are not to be construed as reflecting the views or indorsement of the Department of the Air Force.

population of all shapes or to prescribe proper methods for sampling it.

An attempt was made to avoid this problem by selecting as stimuli a group of shapes assumed to be representative of a limited population, i.e., the kind of shapes which represent terrain features on maps, aerial photographs, and radarscopes. The propriety of extending the results to situations involving real material of these sorts would, of course, still have to be checked empirically. The stimuli chosen were also judged to have a high degree of similarity to one another. It was assumed that the use of a fairly homogeneous group would lead to a greater consistency in the results. The design of the experiment allowed for a statistical check of this second assumption.

In addition to meeting the above criteria, the shapes used also met three other criteria which were judged to be essential: (*a*) they were sufficiently nonmeaningful, or nonrepresentational, that the results could not reasonably be attributed to the differential effects of past experience; (*b*) they could be given a determinate orientation; and (*c*) it had been shown in previous experiments that discrimination among them is difficult when time of exposure is limited (1).

FIG. 1. Examples of the nonsense shapes used as stimuli. The shapes are shown in the arbitrarily selected 0° position.

METHOD

Apparatus.—The stimuli were presented by means of 2 × 2-in. lantern slides and a projector. Each *S* indicated his response by pressing either of two keys marked "S" (Same) and "D" (Different). The projector was equipped with an Alphax shutter which automatically exposed the stimuli for 1 sec. on every trial. The equipment was so constructed that opening the shutter started an electric timer which stopped when either key was pressed. The *S*'s choice was indicated by a signal light, and the response latency was recorded to the nearest .01 sec. Four *S*s were tested simultaneously. The stimuli consisted of one "standard" and ten "comparison" nonsense shapes. The standard and five of the comparison shapes are shown in Fig. 1 as examples.

Subjects.—The *S*s were 256 basic airmen at Lackland Air Force Base. Thirty-two *S*s were assigned at random to each of eight experimental groups.

Procedure.—The stimuli were presented in pairs, each pair consisting of either (*a*) the standard and one of the comparison shapes ("D" trial), or (*b*) two standards ("S" trial). The *S* was instructed to indicate his judgment as quickly as he could by pressing one of two keys marked "S" and "D." On every trial the pair of shapes remained in view for 1 sec. The trials were presented at the rate of approximately four per minute.

The eight angular positions used were: 0°, 30°, 60°, 90°, 180°, 270°, 300°, and 330°. For each of the eight groups of *S*s, the standard shape, always the left-hand member of the pair, appeared in only one orientation. A different orientation of the standard was used for each group. The right-hand member of the pair appeared in the eight different angular orientations for every *S*. Thus, in the entire experiment there were 64 different combinations of orientations of the paired shapes. Since there were ten "D" pairs and one "S" pair, the total number of combinations of pairs and orientations was 704. Each "D" pair was presented three times, and the "S" pair was presented ten times, making a total of 240 "D" trials and 80 "S" trials per *S*. During the course of the 320 randomized trials the *S*s probably acquired some appreciation of the true proportions of "S" and "D" trials. Initially, however, they presumably assumed they would be shown an equal number of each. The effect of such a set would be to produce, by chance, a disproportionate number of "S" responses and a consequent increase in the number of errors on the "D" trials. Since the differences actually obtained were opposite in direction to the kind described above, the possible

effects of such a bias were not considered in the analysis of the results.

Results

The results were analyzed in terms of the accuracy and latency of shape discrimination at angular separations of 0°, 30°, 60°, 90°, 120°, 150°, and 180°. The angular separation was considered positive if the comparison shape was rotated away from the standard shape in a clockwise direction, negative if it was rotated in a counterclockwise direction.

Accuracy.—Figure 2 shows the relation between discrimination accuracy and angular separation, with "S" and "D" trials plotted separately. It can be seen that the effect of orientation was markedly different in the two kinds of trials. On the "S" trials errors increased more or less steadily as the right-hand shape was rotated away from the position of the left-hand shape in either direction. The total range of errors was from about 20% to 40%. On the "D" trials the changes in percentage of errors, while consistent, were not nearly as great, the total range being from about 12% to 15%.

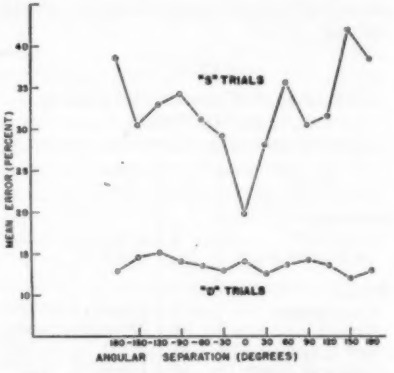


Fig. 2. Accuracy of shape discrimination as a function of angular separation of the stimuli, with "S" and "D" trials plotted separately. Positive angles correspond to a clockwise rotation of the right-hand shape relative to the position of the left-hand shape.

TABLE 1

Analysis of Variance of Errors on "S" Trials

Source	*df*	Mean Square	*F*
Between:			
Standard*	7	24.43	1.07
Ss in same group	248	22.93	
Within:			
Comparison†	7	31.43	12.18**
Standard × Comparison	49	16.86	6.53**
Pooled *S* interactions	1736	2.58	

* Angular position of left-hand shape.
† Angular position of right-hand shape.
** $p < .01$.

Table 1 shows the results of an analysis of variance performed on the data of the "S" trials only. The significant interaction between angular positions of the left-hand shape and angular positions of the right-hand shape ($F = 6.53$, $p < .01$) confirms the impression given by the "S" curve in Fig. 2. It can be asserted that the number of errors on the "S" trials was a function of the relative orientations of the two shapes to be discriminated. A less easily understood result is the significant F associated with the absolute angular position of the right-hand shape ($F = 12.18$), whereas the F associated with the absolute position of the left-hand shape was not significant ($F = 1.07$). This result is probably attributable to the different degrees of precision involved in the two different error terms used in obtaining the F values.[2]

The results of an analysis of variance performed on the data from the

[2] Since differences in angular position of the left-hand shape (Std) were confounded with differences between independently selected groups of *Ss*, these differences were tested by an error term based on *Ss within the same group*. Within-group differences were tested by an error term composed of *pooled subject interactions*.

TABLE 2

ANALYSIS OF VARIANCE OF ERRORS ON "D" TRIALS

Source	*df*	Mean Square	*F*
Between:			
Standard*	7	8.15	1.41
Ss in same group	248	5.80	
Within:			
Comparison†	7	9.50	28.05**
Shape‡	9	114.36	337.64**
Std × Comp.	49	1.10	3.25**
Comp. × Shape	63	3.04	8.98**
Std × Shape	63	.91	2.68**
Std × Comp. × Shape	441	.62	1.83**
Pooled *S* interactions	19593	.34	

* Angular position of standard (left-hand) shape.
† Angular position of comparison (right-hand) shape.
‡ Shape of comparison figure.
** $p < .01$.

"D" trials are shown in Table 2. This analysis involved a three-way classification of the data, since there were ten different comparison (right-hand) shapes. As in the previous analysis, the differences associated with the angular positions of the left-hand shape were not significant; all other main effects and interactions were significant beyond the 1% level of confidence. The appearance of a significant triple interaction term (Std × Comp. × Shape) shows that the curve in Fig. 2 showing the effect of angular separation on discrimination accuracy must be interpreted in terms of the individual shapes used. In other words, the sample of shapes was not homogeneous with respect to the effect of angular separation.

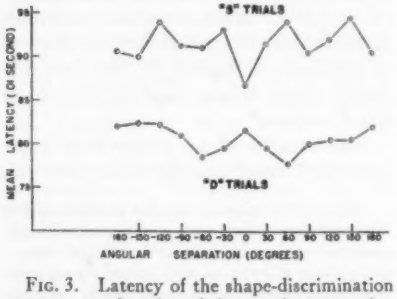


FIG. 3. Latency of the shape-discrimination response as a function of the angular separation of the stimuli, with "S" and "D" trials plotted separately. Positive angles correspond to a clockwise rotation of the right-hand shape relative to the position of the left-hand shape.

TABLE 3

ANALYSIS OF VARIANCE OF LATENCIES ON "S" TRIALS

Source	*df*	Mean Square	*F*
Between:			
Standard*	7	3266.57	1.34
Ss in same group	248	2436.05	
Within:			
Comparison†	7	5989.00	619.33**
Standard × Comparison	49	818.86	84.59**
Pooled *S* interactions	1736	9.67	

* Angular position of left-hand shape.
† Angular position of right-hand shape.
** $p < .01$.

Latency.—The latency data were analyzed in the same way as the error scores. Figure 3 shows mean latency plotted as a function of angular separation. These means were based only on those trials on which the correct response was given. The latency curves bear considerable resemblance to the error curves shown in Fig. 2, particularly with respect to the consistently higher values obtained on the "S" trials.

TABLE 4

ANALYSIS OF VARIANCE OF LATENCIES ON "D" TRIALS

Source	*df*	Mean Square	*F*
Between:			
Standard*	7	25969.57	1.29
Ss in same group	248	20165.32	
Within:			
Comparison†	7	17999.00	75.75**
Shape‡	9	77238.00	325.08**
Std × Comp.	49	532.29	2.24**
Comp. × Shape	63	17630.03	74.20**
Std × Shape	63	595.73	2.51**
Std × Comp. × Shape	441	562.15	2.37**
Pooled *S* interactions	19593	237.60	

* Angular position of standard (left-hand) shape.
† Angular position of comparison (right-hand) shape.
‡ Shape of comparison figure.
** $p < .01$.

The analysis of variance of these data is shown in Tables 3 and 4. It can easily be seen that the results were almost identical to the results obtained in the analysis of errors. Both curves in Fig. 3 deviate significantly from the horizontal (significant Std × Comp. interaction), but the values obtained on the "D" trials also differed significantly as a function of the particular comparison shape used (significant Std × Comp. × Shape interaction). This last result again points up the nonhomogeneity of the sample of shapes with respect to the differential effects of orientation.

Discussion

It was stated earlier that an attempt would be made to avoid the problem of drawing a representative sample from the universe of nonsense shapes by selecting stimuli which were homogeneous with respect to their interaction with the independent variable (angular separation) and, at the same time, representative of a limited population of real shapes. Due to the fact that no psychologically meaningful unit of measurement exists which can be used to describe the attribute of "shape," it was impossible to determine a priori how well the shapes finally selected met these two criteria. It was pointed out that a statistical test of the homogeneity of the shapes was supplied by the design of the experiment, but that the extent to which the shapes represented a larger population (terrain features on maps, photographs, and radarscopes) would have to be determined by future experimental work.

A test of the consistency of the relationship between angular separation and discrimination within the sample of shapes was provided by the triple interaction term of the analysis of the "D" trials. In the case of both errors and latencies this term yielded a significant F, showing that the shapes could not be considered a homogeneous sample. In other words, either the pre-experimental selection of the shapes was not based on those attributes most relevant to the sort of homogeneity desired, or the selection was not sufficiently rigorous. An example of the limitations that this sort of interaction places on the conclusions that can be drawn is given by the following hypothetical experiment. Suppose a psychologist were interested in determining the effect of angular position on the recognition of letters of the alphabet. No problems would be encountered with letters such as **c, e, g, k,** etc., which are clearly different in different angular positions. However, **x** might be rotated through 90° or 180° without the rotation being detectable. Likewise, 180° rotation would make **p** resemble **d** and **b** resemble **q**; **z** and **l** would be unchanged by this amount of rotation, and a perfectly circular **o** looks the same in all positions. The result of these relationships would be that recognizability would be affected by the interaction between angular position and the shape of the particular letter concerned. As a consequence, the hypothetical experimenter would be unable to draw any *general* conclusions about the effect of angular position on the recognizability of letters. Apparently, the same phenomenon (in less exaggerated form) occurred in the present experiment in spite of the effort made to select a homogeneous sample of shapes.

Despite the inhomogeneity of the sample, however, there are some conclusions which appear to be justified. It appears fairly certain, for example, that the characteristically higher scores obtained on the "S" trials would have occurred even though one of the other shapes had been chosen as the standard. Also, an examination of the "D" error curves for the individual shapes suggested that the systematic increase in errors as a function of increasing angular separation is a phenomenon peculiar to the "S" trials only. This is not to say that the discrimination of true shape differences is unaffected by changes in angular separation; rather, it points up the fact that the response has multiple determinants. Two probable determinants are (*a*) the tendency for angular separation to make any two shapes look different, and (*b*) the tendency for angular separation to make more difficult *S*'s task of comparing the two shapes. In the case of identical

shapes, both these factors would operate to increase errors; in the case of different shapes, these factors would tend to work in opposite directions. This explanation would account for the fact that on the "S" trials errors increase with increasing angular separation, while on the "D" trials the number of errors appears to be relatively unaffected.

Even though this experiment failed to provide conclusions about shape perception in general, the results do suggest hypotheses concerning certain practical problems. For example, the results suggest: (*a*) that navigation in flight by means of comparing contours on a map with contours on the ground should be maximally efficient when the map and the ground have the same compass orientation; (*b*) that failure to line up the map and the terrain should lead to more errors in judging "sameness" than in judging "difference"; and (*c*) that the relative preponderance of such "S" errors should increase as a function of increasing angular separation between map and terrain. Verification of these predictions in an actual flying situation would provide clues for methods of eliminating this source of navigational error.

It should be noted that the difficulty encountered in attempting to obtain a suitable sample from the population of two-dimensional shapes is of the same general nature as the methodological problem which has plagued psychologists who have tried to use as independent variables such concepts as *motivation*, *stress*, *permissiveness*, *conceptual element*, and *meaning*. Giving adequate operational definitions to variables such as these enables *E* to draw conclusions generalizable to other samples from the same population of *S*s. He cannot draw conclusions, however, about other values of the same variable. In the case of all the variables listed above, the values used in any given experiment are necessarily a nonrandom selection from a distribution having unknown parameters. The attribute of "shape," although ostensibly and superficially a more point-at-able concept than motivation or stress, actually belongs in the same category and suffers from the same limitations as an independent variable. Until it is possible to specify "shape" in a manner comparable to the specification of size, weight, or area, little progress can be made toward developing a theory of shape perception.

Summary

Pairs of two-dimensional nonsense figures were presented tachistoscopically to *S*s who were required to judge them for identity of shape. Each shape could be rotated clockwise or counterclockwise in its own plane to any of eight angular positions. The combinations of the angular positions of the two shapes in each pair resulted in angular separations of the shapes in multiples of 30° between 0° and 180°, inclusive. The accuracy and latency of shape discrimination were studied as a function of the angular separation of the shapes.

The results may be summarized as follows:

1. On the trials on which the two shapes were actually identical, errors increased as a function of increased angular separation. The range of errors was from 20% to 40%.
2. On the trials on which the two shapes were actually different, errors showed an initial decrease (30° separation), followed by an increase (60°–150°) and a final drop at 180°. The range of errors on these trials was from 12% to 15%.
3. Latencies on the "S" trials were consistently higher than on the "D" trials. The changes in latency as a function of angular separation were statistically significant, though irregular.
4. Generalization of these results is limited by the fact that, in the case of both errors and latencies, a significant interaction was found between angular separation and the variable of shape. The use of "shape" as an independent variable is difficult because it is impossible to specify the procedure for selecting a representative sample of shapes to be used as stimuli.

References

1. Arnoult, M. D. Accuracy of shape discrimination as a function of the range of exposure intervals. *USAF, Hum. Resour. Res. Cent. Res. Bull.*, 1951, No. 51–32.
2. Brunswik, E. Systematic and representative design of psychological experiments. In J. Neyman (Ed.), *Proceedings of the Berkeley symposium on mathematical statistics and probability*. Berkeley, Calif.: Univ. of California Press, 1949. Pp. 143–202.
3. Gibson, J. J., & Robinson, D. Orientation in visual perception; the recognition of familiar plane forms in differing orientations. *Psychol. Monogr.*, 1935, **46**, No. 6 (Whole No. 210).

(Received July 23, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

EFFECT OF STIMULUS AREA AND INTENSITY UPON THE LIGHT-ADAPTED ELECTRORETINOGRAM

JOHN C. ARMINGTON AND FREDERICK C. THIEDE

Army Medical Service Graduate School

The relation of the retinal action potential (electroretinogram or ERG) to the area and intensity of the visual stimulus is of considerable importance (**6, 9**). Study of area and intensity effects may, among other things, provide information regarding glare and stray light effects within the eye (**7, 8**). In addition, attempts may be made to relate electrical findings to the structure of the retina. Of the number of components in the human ERG, some can be identified with specific retinal functions. The prominent *B-wave*, for example, has been shown to possess scotopic properties (**11, 13**), while the *A-wave* has both photopic and scotopic segments (**4, 15**). The *X-wave* accompanies photopic activity (**1, 3**). Area and intensity have been found to be closely interchangeable in their capacity to arouse a *B-wave* of a given magnitude over a wide range of stimulus diameters (**8**), but in this respect photopic potentials (the *X*- and photopic *A-waves*) have received less thorough examination. Thus, some question exists as to whether the density of cone receptor units at the retinal point directly stimulated underlies the sensitivity and magnitude of the *A*- and *X-waves* (**1**). The present experiment, designed to check these possibilities, is primarily a study of the effects of area and intensity upon these two components of the photopic electroretinogram.

METHOD

Apparatus.—The apparatus used in the present study was nearly identical with that employed in earlier investigations (**2, 7, 11**). Light from a tungsten ribbon filament maintained at a constant temperature of 2800° was collimated and passed through interchangeable Wratten neutral density filters and a piece of Corning HR signal 1 red No. 12 glass. To provide single test flashes, the light was brought to a focus upon the leaves of a photographic shutter set at .01 sec. Then the rays diverged and with additional lenses were finally brought to a second focus at *S*'s pupil. In this way, *S* was presented with a "Maxwellian view" of the final lens in the system. The stimulus appeared as a uniform disk of red light.

The diameter of the stimulus (Table 1) was controlled by an adjustable diaphragm placed behind the final lens. A biting board and a dim red fixation point were used to fix the position of the eye. In order to eliminate scotopic effects from the tracings, only red flashes were used, and the eye was maintained in a light-adapted state by an illuminated matte white screen mounted concentrically about the stimulus area. This surround, which nearly filled the visual field, had a luminance of 2 ft.-L.

In the first phase of the experiment, stimuli of different areas were centered on the fovea. In the second phase, a few data were collected with a fixed stimulus of 12° diameter centered at one of several positions on the nasal part of the retina. To stimulate peripheral parts of the retina, the eye was shifted to one side. The actual angular extent of this movement was determined by the fixation point, which was placed at selected locations in the visual field. Actually, stimuli were centered 0°, 12°, 23°, 33°, 41°, and 52° away from the fovea. In the latter phase, a piece of ground glass was mounted in place of the final lens, making alignment with the optical system less critical. In a third phase of the experiment, test flashes were placed in the blind spot, using both of the stimulus arrangements described above.

Potentials were obtained with a silver-silver chloride electrode supported by a contact lens. The reference electrode was worn on the forehead. Recordings were traced with a Grass inkwriting electroencephalograph.

Procedure.—At the beginning of an experimental session, *S* inserted his contact lens electrode and aligned his eye with the stimulator. Next, the eye was adapted to darkness for 15 min. and subsequently adapted to light for 5

TABLE 1

AREAS AND DIAMETERS OF TEST FLASHES

Diameter of Stimulus in Degrees	Estimated Retinal Area (Mm.²)	Log Retinal Area
30	65.8	1.818
26	49.0	1.690
21	31.2	1.494
14.9	17.6	1.246
10.7	7.83	.893
5.3	1.95	.290
2.6	.394	$\bar{1}$.595
1.27	.120	$\bar{1}$.079

min. Experimentation commenced with the adaptation screen remaining lighted. The stimuli were administered at 1-min. intervals, area and intensity being changed in a random manner. Since the contact lens caused considerable discomfort after about one hour's work, as many as ten experimental sessions were sometimes required of *S*. In general, a single response to each area-intensity combination was obtained.

Our records exhibited both *A*- and *X-waves* (**2, 3, 14**). For purposes of measurement, the amplitude or height of the photopic *A-wave* was taken as the distance from the base line immediately preceding the stimulus to the trough of the first negative response; that of the *X-wave*, from this trough to the following crest.

Subjects.—Three *S*s with normal vision were used in these experiments.

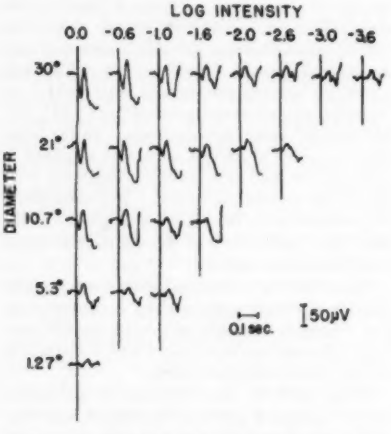


FIG. 1. Typical responses to stimuli of several areas and intensities. Vertical line shows approximate time of the onset of the stimulus. Subject WZ.

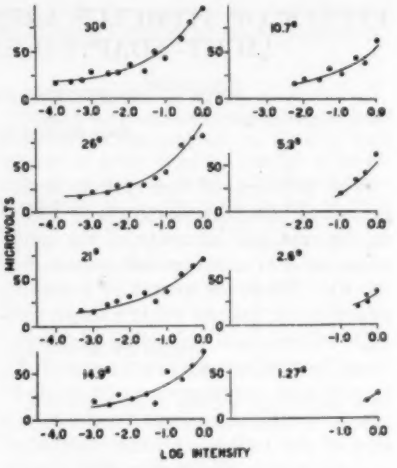


FIG. 2. Height of *X-wave* as a function of stimulus intensity for different areas. Subject FT.

RESULTS

Characteristics of the response.—The features exhibited by our recordings may be seen in Fig. 1, where an upward deflection denotes positivity of the cornea with respect to the reference electrode. Each response displays two principal waves, namely, an initial negative photopic *A-wave* fol-

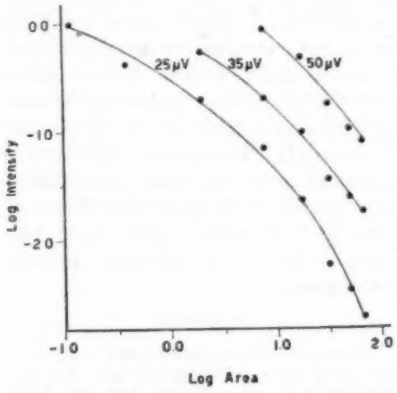


FIG. 3. Relation between stimulus area and intensity for constant heights of *X-wave* response. Average data for three *S*s.

TABLE 2

LOGARITHMS OF INTENSITIES REQUIRED FOR CRITERION HEIGHTS OF *X* RESPONSE

Criterion	*S*	Diameter							
		30°	26°	21°	15°	11°	5.3°	2.6°	1.3°
25 μV	JA	−2.9	−2.7	−2.4	−1.8	−1.0	−0.7	−0.5	0.0
	WZ	−2.8	−2.5	−2.2	−1.5	−1.1	−0.5	− .25	0.0
	FT	−3.0	−2.4	−2.2	−1.9	−1.5	−0.9	−0.4	0.0
	Mean	−2.9	−2.5	−2.3	−1.7	−1.2	−0.7	−0.4	0.0
35 μV	JA	−1.7	−2.0	−1.8	−1.2	−0.6	−0.2		
	WZ	−1.9	−1.6	−1.3	−0.8	−0.6	−0.2		
	FT	−1.8	−1.6	−1.3	−1.0	−0.8	−0.5		
	Mean	−1.8	−1.7	−1.5	−1.0	−0.7	−0.3		
50 μV	JA	−1.2	−1.1	−1.0	−0.4	−0.1			
	WZ	−1.1	−0.9	−0.7	−0.3	−0.1			
	FT	−1.1	−0.9	−0.6	−0.5	−0.2			
	Mean	−1.1	−1.0	−0.8	−0.4	−0.1			

lowed by a positive *X-wave*. With many *S*s, as is the case in Fig. 1, a large additional negative signal follows the *X-wave*. This, however, is largely an artifact potential resulting from movement of the eyelids. Since it is released by the visual stimulus, it must occur subsequent to initial activity of the retina.

It is quite clear that both *A*- and *X-waves* are more prominent with large and intense stimuli: a decrease in either parameter results in lessened activity. Figure 2 presents plots of the amplitude of the *X-wave* response as a function of stimulus intensity for one *S*. The plots, all of the same approximate form, indicate a fairly regular decrease in height with lessened intensity.

From such intensity plots, curves were derived showing the area-intensity combinations which may be taken for a constant *X-wave* height. Figure 3 is based upon such data averaged for all three *S*s; the values for each individual are given in Table 2. These curves again emphasize the effectiveness of area for eliciting the *X-wave*. In fact, it is to be noted that a slight curvature is exhibited by these plots, a feature which demonstrates that *X-waves* were elicited with less total flux when large stimulus patches were used.

Similar plots have been constructed with the larger area data for the photopic *A-wave*, and are shown for one *S* (WZ) in Fig. 4 and 5. Since this component is too small to be reliably measured when small stimulus areas are used, only the responses to large areas are plotted.

Retinal location.—In the second phase of this work, stimuli of fixed

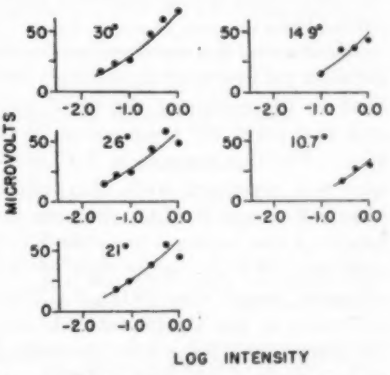


FIG. 4. Height of *A-wave* as a function of stimulus intensity. Subject WZ.

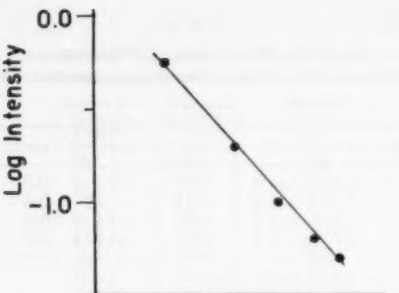


FIG. 5. Relation between stimulus area and intensities used for a 20 μV *A-wave*. Subject WZ.

diameter were placed at different retinal loci. In their form the responses from each position appeared quite similar (Fig. 6), but in size the peripheral areas consistently gave somewhat larger signals than the central ones. Further information on these points was obtained with one *S* (JA) by delivering ten stimuli of maximum intensity to the 0° and 52° positions. The mean amplitude of response for the fovea was 22 μV, while that for the periphery was 35 μV, a difference which was shown to be significant beyond the .01 level of confidence with Wilcoxon's method of unpaired replicates (**16**).

Blind spot.—A comparison was made between the effects of delivering a test flash to the region of the blind spot and to a point 10° temporal to blind spot. For this purpose, a 5.3° stimulus was presented with Maxwellian view. For one *S* (JA), the average height of the response from the blind spot was 38.5 μV, while that of the adjacent region was 39.0 μV. This difference is not significant. A similar comparison made with the ground glass in place also showed no difference.

DISCUSSION

The present results are puzzling because the magnitude of the *A-* and *X-waves* of the ERG does not appear to be directly related to the spatial distribution of cone receptors in the retina. Small stimuli placed where the cones are the most concentrated have no more effect, and sometimes even less effect, than similar stimuli located where cones are sparse. Although the data hitherto available offer no completely adequate explanation for this finding, some of the factors most likely to be involved will be discussed.

Stray light.—The findings reported above closely resemble those found in studies for the *B-wave* (**7, 8**) in that reductions in intensity may be readily compensated by increases in the illuminated area, and responses can be evoked by stimuli

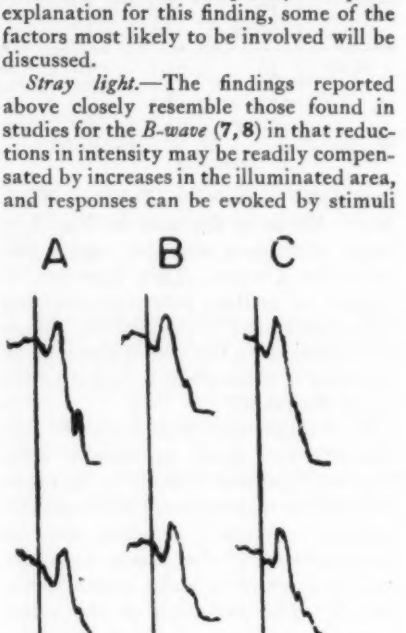


FIG. 6. Responses from six different retinal positions. Subject JA. Locations of stimuli: A, fovea; B, 12°; C, 23°; D, 33°; E, 41°; and F, 52° nasally from the fovea.

focused upon the blind spot (5, 7, 8). To account for this, the hypothesis has been advanced that the *B-wave* is not produced exclusively by those receptors directly stimulated, but, rather, that it is released by the entire scotopic retinal population indirectly stimulated through entopic stray light. A simple stray light hypothesis, however, does not adequately explain *X*- and photopic *A-wave* activity. A large segment of the photopic mechanism must have been directly stimulated with flashes of all areas since the cone receptors are greatly concentrated in the region of the fovea and its immediate surroundings. Because of this and because less of the retina remains to be indirectly stimulated as stimulus area is increased, it would be anticipated that as area became greater, additional increments would become less rather than more effective.

There is, nevertheless, reason for believing that stray light has some effect on the components of the photopic ERG. Otherwise, it is difficult to explain the large response to stimuli centered about the blind spot. Furthermore it has been reported that the photopic ERG is quite sensitive at short wave lengths (3); this phenomenon may also be attributed, in part, to stray light effects.

Origin of the ERG.—There is little evidence regarding the site from which the contact lens electrode collects potentials, but located as it is at the front of the eye, it may be exposed mostly to peripheral retinal activity. If this should be true, a major part of the recorded response would be produced at the periphery by scattered light, while response directly produced by more central regions of the retina would be less conspicuous. It is interesting to note in this connection that if a voltage is applied to our electrodes so as to elicit phosphenes, these appear to be confined to peripheral regions. A second factor to be considered is that we do not know which structures are responsible for the major features of the ERG. This is especially true of the vertebrate eye.[1]

[1] Evidence of Hartline and his associates (10) suggests rather strongly that the ERG of Limulus originates in the receptors.

The fact that large responses can be obtained from peripheral regions of the retina where the retinal population is relatively sparse might indicate that neural elements other than actual receptors are involved in the production of the ERG. Regional differences in retinal tissue volume are less pronounced than is the density of the photopic receptor population (12). Thus, the possibility must be kept in mind that, by virtue of a greater wealth of nervous connections, peripherally situated receptors might be capable of stimulating a comparatively large number of bipolar and ganglion cells.

A-wave.—A straight line has been drawn through our area-intensity wave for the *A-wave* (Fig. 5). The exact form of this curve, which covers a limited range of intensities, is not certain. *A-waves* of 20 μV, however, appear only when the *X-wave* is quite large, and examination of Fig. 3 will reveal that the 50 μV *X-wave* data could also be well described by a straight line of the same slope. Hence, the results do not point out differences, but on the contrary, point to the similarity of the two waves.

Summary

Experiments were conducted to test the effects of stimulus area and intensity upon the light-adapted electroretinogram. Conventional recording procedures and techniques were used. It was found that inversely related stimulus areas and intensities may be used to produce criterion magnitudes of both *A*- and *X-wave* responses. The *A*- and *X-wave* responses do not appear to depend in any simple way upon the retinal region which is directly stimulated. The possibility that scattered light, or other factors, may be responsible for the results is discussed.

References

1. Adrian, E. D. The electrical responses of the human eye. *J. Physiol.*, 1945, **104**, 89–104.
2. Armington, J. C. A component of the human electroretinogram associated with red color vision. *J. opt. Soc. Amer.*, 1952, **42**, 393–401.
3. Armington, J. C. Electrical responses of the light-adapted eye. *J. opt. Soc. Amer.*, 1953, **43**, 450–456.

4. Armington, J. C., Johnson, E. P., & Riggs, L. A. The scotopic A-wave in the electrical response of the human retina. *J. Physiol.*, 1952, **118**, 289–298.
5. Asher, H. The electroretinogram of the blind spot. *J. Physiol.*, 1951, **112**, 40.
6. Bartley, S. H. *Vision: a study of its basis.* New York: Van Nostrand, 1941.
7. Boynton, R. M. Stray light and the human electroretinogram. *J. opt. Soc. Amer.*, 1953, **43**, 442–449.
8. Boynton, R. M., & Riggs, L. A. The effect of stimulus area and intensity upon the human retinal response. *J. exp. Psychol.*, 1951, **42**, 217–226.
9. Granit, R., Rubinstein, B., & Therman, P. O. A new type of interaction experiment with the retinal action potential. *J. Physiol.*, 1935, **85**, 34–36.
10. Hartline, H. K., Wagner, H. G., & MacNichol, E. F., Jr. The peripheral origin of nervous activity in the visual system. *Cold Spr. Harb. Sympos. Quant. Biol.*, 1952, **17**, 125–141.
11. Johnson, E. P. The electrical response of the human retina during dark adaptation. *J. exp. Psychol.*, 1949, **39**, 592–609.
12. Polyak, S. L. *The retina.* Chicago: Univ. Chicago Press, 1941.
13. Riggs, L. A., Berry, R. N., & Wayner, M. A comparison of electrical and psychophysical determinations of the spectral sensitivity of the human eye. *J. opt. Soc. Amer.*, 1949, **39**, 427–436.
14. Schubert, G., & Bornschein, H. Beitrag zur Analyse des menschlichen Elektroretinogrammes. *Ophthalmologica*, 1952, **123**, 396–413.
15. Vukovich, V. Das ERG des Achromaten. *Ophthalmologica*, 1952, **124**, 354–359.
16. Wilcoxon, F. Probability tables for individual comparisons by ranking methods. *Biometrics*, 1947, **3**, 119–122.

(Received August 3, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

PERCEPTION OF THE RELATIVE DISTANCE POSITION OF OBJECTS AS A FUNCTION OF OTHER OBJECTS IN THE FIELD[1]

WALTER C. GOGEL

Army Medical Research Laboratory

It is well known that a number of circumstances, such as the presence of misleading cues of size, may distort *O*'s perception of the relative distance of objects from himself. The purpose of the present study is to determine what effect such a distortion will have upon the perception of the distance position of objects other than those that contain the misleading cues.

The nature of this problem can be illustrated by a landscape containing a large and small building, with the larger building more distant than the smaller. Assume that cues which would make *O* perceive the actual distance of the buildings are reduced or absent so that the more distant building, because of its larger size, appears nearer than the smaller building even though it is physically farther away. In this case, what would be the perceived relative distance of a tree that is located beside the apparently nearer (larger) building? If, for example, the tree is correctly perceived as being at the distance of the apparently nearer building, it will be incorrectly perceived as being less distant than the apparently farther (smaller) building. If, however, the tree is correctly perceived as being more distant than the apparently farther building, it will be incorrectly perceived as also being more distant than the apparently nearer building. If the tree appears intermediate in distance to the two buildings, its distance will not be correctly perceived with regard to either building. Also, what would be the perceived distance of the tree in relation to other objects, for example, a bush which is physically located beside the apparently farther building? If such questions are asked with respect to various objects scattered throughout the landscape, it will be seen that the distortion in the perception of the relative distance of the two buildings will probably affect the perception of the relative distance of many other objects, both as compared with the distance to the buildings and to each other.

Visual situations of this type have two essential characteristics. First, a difference between the physical and apparent relative distance of several objects (an illusion in depth) must be present. Second, an object which does not involve the misleading cues to depth (a dissimilar object) must be located in the visual field. The study of such visual situations requires that the apparent distance of the dissimilar object with respect to the various parts of the visual field be investigated.

In the experiments that follow, the illusion was produced by a series of apparently rectangular figures which differed in size. The dissimilar object was a disc. Since the disc differed in shape from the rectangles, it presumably would not result in relative size cues with respect to these rectangles. The investigation of the apparent distance of the disc was accomplished through determining the physical positions at which the disc appeared equidistant with several parts of the illusion. Short viewing distances of approximately 10 ft. were used, rather than the larger distances which might be suggested by the example of the landscape.

General Apparatus

The disc, rather than being an actual object, was generated by a stereoscopic instrument which projected a disc of light on the retina of each eye of *S*. Two independent optical systems were used, one for producing each disc of light. Each optical system consisted essentially of a concave (Mangin) mirror which was placed

[1] The author wishes to acknowledge the valuable assistance of John P. Tammaro, David A. Nordling, and Fred L. Nungesser, Jr. in making the series of figures and in collecting and analyzing the data.

above a circular aperture source of light so that parallel light was reflected from the mirror. The light coming from the mirror was reflected from a half-reflecting surface into the eye of *S*. The half-reflecting surface permitted *S* to see both the disc and actual objects which were placed a distance in front of *S*. When *S* turned a knob on the stereoscopic instrument, the right optical system rotated and changed the lateral position of the disc of light on the right retina. The left optical system always remained fixed. The disc was orange in color and subtended 15 min. of visual angle.

When *S* looked through the stereoscopic instrument binocularly, the two retinal disc images fused and a single disc was seen. If another object, for example, a rectangle, was physically located in front of *S*, the disc appeared to be in front of, behind, or at the same distance as the rectangle, depending upon the knob adjustment. The knob controlled the retinal disparity between the fused disc and the rectangle, so that turning the knob made the disc appear to move in distance. The stereoscopic instrument was so mounted that the whole instrument could be rotated. This rotation did not change the retinal disparity relations between the disc and any object in front of *S*, but it did change the lateral (right-left) position of the disc in the visual field. The adjustment knob on the stereoscopic instrument was linked to an indicator which measured the orientation of the right optical system. With *S*'s interpupillary distance (as measured with an interpupillometer) and the calibration constant of the instrument, the linear distance which corresponded to the position of the disc could be calculated for any adjustment of the knob. The instrument used in this study has been described by Alluisi and Harker (1), and a later model by Harker (3).

Experiment I

Apparatus and Procedure

In Exp. I, the illusion in depth was produced by using a series of nine white similar figures. These figures (trapezoids) varied regularly in size. They were made from a piece of white cardboard, parts of which were blackened so that the remaining parts formed the figures. The over-all dimensions of this series of nine figures were 7.4 cm. in height at the larger end, 2.5 cm. in height at the smaller end, and 40 cm. in length.[2]

When *S* looked through the stereoscopic instrument, he saw the disc and what appeared to be a series of nine white rectangles. In what is called the "illusion situation," the rectangles seemed to be in a row which extended from near right to far left. Actually, as shown in Fig. 1, the right rectangles were farther from *S* than the left rectangles. If the position of *S* were indicated in the two top views of Fig. 1, it would be well below the bottom of the drawings. When the same series of figures was orientated as shown in the right half of Fig. 1, the right rectangles appeared to be farther away than the left rectangles. This is in agreement with their physical positions and is called the "nonillusion" situation. In both the illusion and nonillusion situation, the extreme right and extreme left edges of the display were 322 cm. and 284 cm., respectively, from *S*. In order to determine how the illusion affected the apparent distance position of the disc, *S* adjusted the disc to apparent distance equality with several parts of the illusion. Repeating this in the nonillusion situation permits the effects due to the illusion to be evaluated.

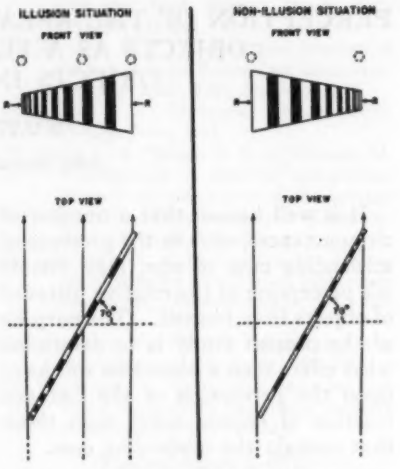


Fig. 1. Schematic drawing of front and top views of the illusion and nonillusion orientation of the series of figures, showing the various lateral positions of the disc and the two positions of the reference object (R)

The *E* rotated the stereoscopic instrument so that the disc appeared in one of three lateral positions. These lateral positions were not changed by *S*. The three alternate lateral positions of the disc are indicated by the small dashed circles in the front view of the left half of Fig. 1. The lateral position of the disc determined its line-of-sight position with respect to the series of figures, but did not determine the apparent dis-

[2] Additional information on the construction of the series of nine figures and the control rectangles can be found in Army Medical Research Laboratory Report No. 107, Fort Knox, Kentucky, 1953.

tance of the disc from *S*. This is illustrated by the long arrowed lines in the top views of Fig. 1. For example, for all points on the left long arrowed line in the left half of Fig. 1, the disc was most nearly in line of sight with the left end of the series of figures. For purposes of description, the disc in this case will be said to be "laterally adjacent" to the left end of the series of figures, but "laterally displaced" from the right end of the series of figures. These long arrowed lines also represent the approximate path along which the disc could be moved in distance by *S* when it was in each of the three lateral positions. For example, when the disc was laterally adjacent to the left end of the series of figures, it remained laterally adjacent to this end, regardless of whether *S* adjusted the knob to make the disc appear in front of, behind, or at the distance of this left end. As shown in Fig. 1, all three lateral positions of the disc were used with the illusion situation, but only the two extreme lateral positions were used with the nonillusion situation.

With the disc in each of the three lateral positions, *S* was instructed to fixate either the extreme right or the extreme left edge (end) of the series of nine figures, and to adjust the disc in depth until it appeared at the same distance from himself as this edge. The edge selected for this purpose is called the reference object. The two alternate positions of the reference object are shown in each of the top views of Fig. 1. It will be seen from Fig. 1 that the lateral position of the disc was either adjacent to the reference object or displaced from the reference object to either the opposite end or the approximate center of the series of nine figures. The use of the series of nine figures is called an experimental situation.

Errors in adjusting the disc to equidistance with a reference object might be due to (*a*) the presence of the series of nine differently sized figures, (*b*) errors in the measurement of the interpupillary distance of *S* or in the measurement of the calibration constant of the stereoscopic instrument, or (*c*) psychological errors attributable to sources other than the presence of the series of figures; for example, aniseikonic effects when the disc was laterally displaced from the reference object. Only errors attributable to *a* are of interest in this study. Therefore, control situations were used in which the only errors which could occur were from sources *b* or *c*. The differences between the results obtained in an experimental situation and its appropriate control can be attributed to the presence of the series of nine figures. A control situation consisted of a single small rectangle with the series of nine figures removed. The rectangle was placed frontal to *S*, with its outer edge at the spatial position previously (or later) occupied by a reference object of the series of nine figures. In this position, a control rectangle had approximately the same retinal size as that end figure of the series which contained the reference object for which it was a control. This required four control rectangles which were presented one at a time. With the illusion situation, the control rectangle for the left reference object was a small rectangle located at the left at a distance of 284 cm. from *S*, and for the right reference object, a larger rectangle located at the right at a distance of 322 cm. from *S*. With the nonillusion situation, the control rectangle for the right reference object was a small rectangle located at the right at a distance of 322 cm. from *S*, and for the left reference object, a larger rectangle located at the left at a distance of 284 cm. from *S*. The outer edge of a control rectangle was used as a reference object. The lateral (and vertical) position of the disc with respect to a particular reference object was the same in both an experimental and its appropriate control situation. All the objects, including the disc, were seen binocularly throughout this experiment.

When the disc was laterally adjacent to a control rectangle, errors due to source *c* were minimal. The adjustment of the disc to equidistance with a control rectangle, in these cases, can be considered to be correct; i.e., all the error can be attributed to source *b*. Laterally displacing the disc from a control rectangle in these experiments resulted in only small changes in the distance adjustment of the disc. Therefore, the average adjustments from all the control situations can be considered as approximately correct.

Twelve men were used as *S*s. Each *S* made 16 adjustments of the disc to apparent depth equality with each reference object in each experimental and control situation. Following a set of 16 adjustments in an experimental situation, the *S*s were asked to estimate the orientation of the series of nine figures. The order of presenting the different situations was systematically varied between *S*s with respect to (*a*) whether a control was presented before or after the experimental situation to which it applied, (*b*) whether the illusion or nonillusion situation was presented first, (*c*) the order of presenting the three lateral positions of the disc with respect to the reference object, and (*d*) the order of designating the right or left edge as a reference object.

In the three experiments of this paper, the brightness of the objects actually located in front of *S* was 2.4 ft.-lamberts; the illumination was white, and no objects were visible except the disc and either the series of nine figures or a control rectangle.

Results

The results from Exp. I are summarized in Table 1. Each mean in

TABLE 1

MEANS AND *SD*'S IN CENTIMETERS FROM ADJUSTING A DISC TO APPARENT EQUIDISTANCE WITH A REFERENCE OBJECT WHEN THE LATTER IS PART OF A SERIES OF SIMILAR FIGURES OF DIFFERENT SIZES AT DIFFERENT DISTANCES

Lateral Position of Disc	Position of Reference Object	Illusion Situation				Nonillusion Situation			
		Experimental		Control		Experimental		Control	
		Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
At Right End	Right End	328	3.9	328	3.3	331	7.1	331	5.8
At Left End	Left End	294	4.8	294	5.0	294	3.7	295	4.1
At Left End	Right End	288	13.3	330	13.9	324	21.5	330	11.1
At Right End	Left End	336	12.3	293	8.5	299	5.3	290	6.0

the table is an average of 12 scores, one from each *S*, where each score is an average of 16 adjustments of the disc to apparent distance equality with a reference object. The *SD*'s in Table 1 were computed from the distributions of 12 scores. The positions of the disc and reference object given in the first two columns of Table 1 specify the lateral relation between the disc and reference object. For example, with the reference object the right end of the series of nine figures, the first row refers to the situation in which the disc was laterally adjacent to this reference object and the third row to the situation in which the disc was laterally displaced from this reference object to the opposite end of the series of figures. The average adjustments from the situation in which the disc was laterally adjacent to the approximate center of the illusion are not shown in Table 1. These results are 315 cm. ($SD = 5.7$) and 313 cm. ($SD = 8.7$) for the adjustment to the right and left reference objects, respectively.

The average estimation of the orientation of the series of nine figures was that it appeared to extend from near right to far left at 23° with respect to the frontal plane in the illusion situation, and from near left to far right at 43° in the nonillusion situation. Actually, the orientation was always at 70° from near left to far right. It seems that some amount of illusion remained in the nonillusion situation.

The average adjustments in centimeters from the control situations in this and the following experiment are larger than the physical distances of the control rectangles from *S*. A constant error probably occurred in converting the indicator readings of the stereoscopic instrument to centimeters. This is not important, however, since the concern is with the differences between the mean results from the various situations.

Consider the case in which the disc was laterally adjacent to the left end of the series of figures in the illusion situation. When the disc was adjusted to the apparent distance of this left end, the mean adjustment was the same as the mean adjustment on the control (294 cm.), i.e., the adjustment to the left end was correct. In adjusting the disc to the apparent distance of the apparently nearer but physically farther right end, the disc was both physically and apparently moved in front of the left end to a distance of 288 cm. from *S*. It may be concluded that with the disc at this left lateral position, the physical distance of the disc with respect to the left end was always correctly perceived. But

the physical distance of the disc was incorrectly perceived in relation to the right end of the illusion. This is indicated by the difference between the 288-cm. mean adjustment on the right end and the 330-cm. mean adjustment on the identically placed control for this right end. This difference of 42 cm. was significant beyond the .001 level of confidence ($t = 14.6$).

In the case in which the disc was laterally adjacent to the right end of the series of figures in the illusion situation, it can be concluded that the distance position of the disc was always correctly perceived with respect to the right end and always incorrectly perceived with respect to the left end. This is indicated by the fact that in this situation, the average adjustment of the disc to apparent equidistance with the right end of the illusion was the same as when the control was used (328 cm.), but in adjusting the disc to the apparent distance of the apparently farther but physically closer left end, the disc was moved behind the 328-cm. position to a distance of 336 cm. This mean adjustment of 336 cm. was significantly different from that from the control (293 cm.) beyond the .001 level of confidence ($t = 12.1$).

Some differences between the average adjustment of the disc in the experimental and appropriate control situations also occurred with the nonillusion situation when the disc was laterally displaced from the reference object to the other end of the series of figures. These differences were less, however, than those which occurred in the illusion situation (beyond the .001 level of confidence, $t = 7.2$ and 8.4) and can be attributed to *S*'s seeing the so-called nonillusion series at a 43° instead of its 70° orientation. Also, laterally displacing the disc to the approximate center of the series of nine figures in the illusion situation affected the apparent depth relation between the disc and both reference objects. But this effect was less than that which occurred when the disc was laterally displaced from the reference object to the opposite end of the illusion. Evidently, when the disc is laterally adjacent to one object, the distortion in the perception of its distance position with respect to another object depends upon the magnitude of the depth illusion between these two objects.

Experiment II

Apparatus and Procedure

The illusion in depth used in this experiment was similar to that of Exp. I except that the series of figures was parallel to the objective frontal plane of *S*, with each end at a distance of 303 cm. With this orientation, the series of figures was constructed to produce the same retinal image as that which occurred in the illusion situation of Exp. I. The procedure was the same as that of Exp. I with the same 12 *S*s being used, but unlike Exp. I, a nonillusion situation was not presented and only the two extreme lateral positions of the disc were used. The control situations consisted of either a small rectangle located to the left or a larger rectangle located to the right, with each rectangle at a distance of 303 cm. from *S*.

Results

The average estimate was that the series of nine figures extended from near right to far left at 20° with respect to the frontal plane. The results from adjusting the disc to apparent equidistance with the extreme right or the extreme left edge of the series of figures are summarized by the means and *SD*'s of Table 2. The mean adjustment differences between the experimental and appropriate control situations were small or zero when the disc was laterally adjacent to the reference object, but larger when the disc was laterally displaced from the reference object to the other end of the illusion. Neither of the former differences was significant at the .1

TABLE 2

MEANS AND *SD*'S IN CENTIMETERS FROM ADJUSTING A DISC TO APPARENT EQUIDISTANCE WITH A REFERENCE OBJECT WHEN THE LATTER IS PART OF A SERIES OF SIMILAR FIGURES OF DIFFERENT SIZES AT THE SAME DISTANCE

Lateral Position of Disc	Position of Reference Object	Illusion Situation			
		Experimental		Control	
		Mean	*SD*	Mean	*SD*
At Right End	Right End	312	3.0	312	3.2
At Left End	Left End	312	2.6	313	2.6
At Left End	Right End	306	7.5	310	6.9
At Right End	Left End	319	8.5	310	8.9

level of confidence. The latter differences (319 cm. minus 310 cm., and 310 cm. minus 306 cm.) were significant beyond the .02 and .05 levels of confidence (t = 2.7 and 2.3). Like the results of Exp. I, these results indicate that the illusion did not disturb the perception of the distance position of the disc with respect to the part of the illusion with which it was laterally adjacent, but this perception was disturbed with respect to the part of the illusion from which the disc was most laterally displaced. It was as though the binocular disparity between the disc and the part of the illusion from which the disc was most laterally displaced was ineffective in determining the apparent position of the disc. This is investigated further in Exp. III.

EXPERIMENT III

Apparatus and Procedure

The only situation used in this experiment was the illusion situation of Exp. I. No controls were employed, and the disc was always laterally adjacent to the left end, with the extreme right edge of the series of figures always designated as the reference object. In Exp. I and II all the objects were seen binocularly. In this experiment with one situation called the "monocular illusion" only the right (largest figure) and the disc were seen binocularly. The eight other figures of the series of nine figures were seen only with the right eye. In another situation called the "binocular illusion," all the objects including the disc were seen binocularly. Ten *S*s were used, eight of whom were *S*s in Exp. I and II. Each *S* made 16 adjustments of the disc to apparent distance equality with the fixated reference object in both the monocular and binocular illusion situation. With five *S*s (four of whom were used in Exp. I and II), the monocular illusion was presented before the binocular illusion. With the remaining five *S*s the order of presentation was reversed.

Results

The average estimate was that the series of figures extended from near right to far left at 22° with respect to the frontal plane in the monocular illusion, and at 20° in the binocular illusion. Physically the orientation was from near left to far right at 70° with respect to the frontal plane.

With the monocular illusion, binocular disparity did not occur between the disc and those portions of the illusion which were seen with only one eye. The mean adjustment of the disc to apparent equidistance with the right end of the illusion was 317 cm. (SD = 13.5) in the monocular and 291 cm. (SD = 8.5) in the binocular illusion situation. The difference of 26 cm. between these two means is significant beyond the .001 level of confidence (t = 6.7). The disc had to be adjusted physically closer to *S* in the binocular than in the monocular illusion in order to appear at the same distance as the same reference object. It seems that the apparent distance position of the disc with respect to this reference object was determined by the binocular disparities between the disc and objects laterally adjacent to the disc, rather than by the binocular disparity between the disc and reference object.

Discussion

These experiments suggest the following hypothesis: With a binocular illusion in depth, the distance position of a disc will be most correctly perceived with respect to that part of the binocular illusion with which it is most frontally adjacent. The word *frontally* is used rather than *laterally* in order to include situations involving either or both lateral and vertical displacement. The part of the illusion to which the disc is most frontally adjacent is defined as that part which is most visually adjacent to a line that extends from *S* through the disc. It is the object most adjacent to the disc when all the parts of the illusion are projected on the same objective frontoparallel plane from the position of *S*.

Several of the Ames demonstrations (4) provide situations of the general type considered here. Ames (2), using a trapezoidal window, found that an object attached to the physically rotating window would appear to rotate around the window while the window appeared to oscillate. Kilpatrick (5) attached a playing card to the rotating trapezoidal window and found with monocular observation that the card appeared to rotate independently of the window or to oscillate with the window, depending upon the condition of attachment. From the present study, it would be expected that, with binocular observation, the apparent path of movement of the attached object would depend upon its successive line-of-sight (lateral) relations with the various parts of the trapezoidal window. These lateral relations can be affected by the manner in which the object is attached to the window. Therefore, the apparent path of the object should be predictable for particular conditions of attachment. This will be considered in a future report.

The results from the present study suggest a convenient method of investigating the apparent depth between two binocular objects. If a disc (real or stereoscopically generated) is placed nearly in line of sight with one of the two objects, and, while maintaining this position, is successively adjusted to apparent distance equality with each of the two objects, the difference in the two distance adjustments will indicate the apparent depth between the two objects. This would not be the case, however, if the disc were placed so that it was nearly in line of sight with each object before the apparent distance equality adjustment to that object was made.

Also, from the results, it would be expected that if two discs had been present in the illusion situation of Exp. I, with a disc placed above and at the distance of each end of the illusion, the right disc would have tended to appear closer to *S* than the left disc, even though it was physically farther away. This suggests that the apparent orientation of the visual field surrounding the series of nine figures should undergo as much distortion as that which occurred with the series of nine figures. Conversely, if the series of nine figures fails to distort the depth relations in the surrounding visual field, it is likely that the correct orientation of the series of nine figures will be perceived. This might explain why a binocular illusion in depth seems to occur most readily in an otherwise dark room, i.e., in a situation which does not contain a number of other objects whose relative distance positions with respect to each other are clearly perceived.

Summary

An illusion in depth was produced by using a series of similar objects of different sizes. It was found that the adjustment of a binocular disc to apparent distance equality with a part of the binocular illusion was influenced by the line-of-sight position of the disc with respect to the illusion. It is hypothesized that a binocular depth illusion will least disturb the apparent relative distance of the disc from *S* with respect to that part of the illusion which is most nearly in line of sight with the disc. It is suggested that the binocular disparity between the disc and a frontally adjacent (line-of-sight) object is functionally more important than that between the disc and a frontally displaced object, regardless of which of these binocular disparity relations *S* attempts to use.

References

1. Alluisi, E. A., & Harker, G. S. The stereoptometer—a simple haploscopic instrument for the study of binocular space perception. *Science*, 1953, 117, 682–683.
2. Ames, A., Jr. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, 65, No. 7 (Whole No. 324).
3. Harker, G. S. The stereoptometer—an instrument for the study of binocular vision. *Army Med. Res. Lab.* Rep. No. 106, Fort Knox, Kentucky, 1952.
4. Ittelson, W. H. (Ed.) *The Ames demonstrations in perception.* Princeton: Princeton Univ. Press, 1952.
5. Kilpatrick, F. P. Assumptions and perception: three experiments. In F. P. Kilpatrick (Ed.), *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952.

(Received August 6, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

THE EFFECTS OF PUNISHMENT UPON SYLLABLE RECOGNITION THRESHOLDS[1]

WILLIAM LYSAK[2]

University of Wisconsin

Discrepancies in the results of recent experiments have brought into question the concepts of "perceptual defense" and "subception" (1, 2, 3, 5, 6, 7, 8, 9, 11). Some have shown higher recognition thresholds for negative or taboo words as well as greater GSR's during "prerecognition exposures" of such words (7). Others have shown greater GSR's in the prerecognition exposures of previously punished words but no differences in recognition thresholds (5, 6); while another (2) has shown lowered recognition thresholds to both positive and negative value words.

An examination of the conditions under which these experiments were performed suggests that four variables, primarily methodological in character, might account for the discrepancies in results. These are: (*a*) the punishment and/or threat of punishment present when recognition thresholds are obtained; (*b*) the discriminability of the items in the response pool with respect to their affective character; (*c*) the freedom of *S*s to respond or not to respond after each tachistoscopic exposure; (*d*) the extent of the response pool available to *S*. The present series of experiments tests the effects of these variables upon syllable recognition thresholds.

Experiment I

Method

This experiment was designed in the form of $3 \times 2 \times 2$ factorial design to provide information relevant to the first three variables outlined above. The first variable, punishment, was studied under three conditions: punishment, no punishment, and ambiguous threat. Discriminability was investigated at two levels: clear discriminability and moderate discriminability. Freedom to respond was also studied at two levels: free choice and forced choice.

Subjects.—The *S*s were 144 students from undergraduate courses at the University of Wisconsin. They were recruited on a voluntary basis and assigned to the 12 experimental groups at random.

Apparatus.—The apparatus consisted of a University of Missouri model memory drum, a shocking apparatus, a standard Harvard tachistoscope made by Gerbrands, and multicolored reference lists of syllables on which *S*s recorded their choices.

The memory drum contained two sets of the following disyllables, half in red and half in green:

BAHUF	BUXAP
CEROG	CIBUC
DISEH	DECIW
FOTIJ	FAYEL

Syllable colors were reversed on each list. These syllables were selected to be of sufficient length to preclude recognition on the first or second exposure and so that there would be some structural similarity between the corresponding sets of four. Each of the eight syllables appeared twice in random order on the two lists.

The shocking apparatus consisted of a telegraph key connected to a Harvard inductorium which permitted continuous shock changes, a 6-v. battery, and an electrode capable of being affixed to *S*'s ankle.

In order to obviate too early recognition, a Polaroid filter which decreased the transmitted light by half was inserted in the eyepiece of the tachistoscope. A white card containing a small rectangle outlined in black was used for the fixa-

[1] This report is based upon a dissertation submitted to the Graduate School of the University of Wisconsin in partial fulfillment of the requirements for the Ph.D. degree. The author acknowledges his appreciation to Professor J. C. Gilchrist whose stimulating instruction, encouragement, and rigorous criticism made the investigation possible.

[2] Now at William Beaumont Army Hospital, El Paso, Texas.

tion area of the pre-exposure field. Uniform test cards were constructed by having the syllables typed in capitals on an electric typewriter.

Two wooden compartments in the drawer of a table on which the tachistoscope rested contained a supply of reference slips arranged so that *S* dealt with only one, the top, slip at a time. Each slip contained the eight syllables in random order, and sets of slips were constructed so that each syllable appeared once in each position. The slips from these sets were inserted randomly in the compartments. Some of these reference slips were printed in red and green and others only in blue. Those used depended upon the condition under test. Where syllables were to be discriminably different, only those slips containing the syllables in their original red and green colors were available to *S*. Where differences between syllables were not to be clearly discriminable, reference slips showing all the syllables in blue were used.

Procedure.—Upon entering the laboratory, each *S* underwent a brief visual acuity check. He was required to have 20/20 correctible vision and also had to be able to discriminate between red and green. The *S* was then seated between two tables upon which rested the memory drum and tachistoscope.

The *S* was told that he was participating in a syllable perception experiment and that it would be necessary to undergo a slight electric shock. He was instructed to depress a telegraph key for approximately .5 sec. while *E* gradually increased the intensity until *S* indicated his shock threshold. After he had done so, he was told *E* would continue slowly to increase the intensity while *S* continued depressing the key until the shock became painful. The *E* then established the shock at a level halfway between the threshold and the point of pain. This was considered a moderate shock. When a strong shock was needed for a later phase of the experiment, the painful level was utilized.

After the shock levels had been recorded, *S* was asked to attend to the memory drum and was given the following instructions: "Your job is to look at the syllables which appear in this aperture and give yourself a shock by pressing the key every time you see a red (or green) syllable. Press it the way you did before so that the buzz can be heard clearly. Do you understand? Shock yourself every time you see a red syllable. Meanwhile try to learn what the syllables are and which ones get shocked and which ones do not. It is not necessary to try to learn the sequence in which the syllables appear. Learn them in any way you like so as to be able to spell the syllables later. Are there any questions?"

The *S* was also asked to report whenever a trial was concluded. He viewed one list of 16 syllables for each revolution of the memory drum cylinder. Each syllable was visible for 4 sec., and 1 sec. elapsed between syllables. A 20-sec. rest interval occurred between trials. Each *S* was exposed to five revolutions of the memory drum. Thus, he viewed each of the eight syllables ten times.

After completing the familiarization trials, *S* turned to the tachistoscope while *E* placed the shocking key behind a screen away from *S*. All *S*s were then given the following directions: "Get up as close to the eyepiece as you comfortably can. The rectangle you see will disappear and one of the syllables you have been viewing will appear there very briefly. Each syllable will be black rather than red or green and each may appear more than once. After each exposure you are to look at the top slip on this supply of slips and make a check mark next to the syllable you think has been exposed. Then spell out the syllable, turn the slip over and place it in this other compartment. Do the same for successive trials. Are there any questions?"

The following specific instructions were given *S* whenever appropriate.

For the punishment condition *S*s were told, "Now pay attention to this. If the syllable you choose is a shock syllable you will be shocked again after you have made your choice by checking it off and spelling it out. But this time the shock will be much stronger than before so make your choices carefully. The increased shock will be demonstrated. (*E* does so.) Remember you will not be shocked for being either right or wrong. You will be shocked every time you choose a syllable on which you shocked yourself before. Are there any questions?" For this condition the shocking apparatus was set to the previously established pain threshold.

For the no-punishment condition, *S* was instructed to remove the electrode after completing the familiarization trials, and he was told that there would be no additional shock.

For the threat condition, *S* retained the electrode and *E* said only, "O.K. Now, I'll handle the shocking key."

Under the forced-choice conditions, *S* was instructed to choose a syllable after each exposure. When choices were not forced, he was told to check a blank at the bottom of the reference slip for each exposure until the syllable became recognizable and then to check the correct syllable.

For convenience, the test cards were successively rotated and alternated between shock and nonshock syllables for each *S*. (Only 2 of the 144 *S*s were able to detect the alternation when asked about it afterward.) One slip was checked for each exposure until the syllable being shown

was correctly identified. These reference slips constituted the primary data of the experiment. Recognition thresholds were obtained by counting the number of slips accumulated for each test card exposed. Test cards were exposed in an ascending series starting with .01 sec., and each successive exposure was increased in duration by .01 sec. After a threshold had been obtained for each syllable, *S* was asked to recall as many syllables as could be remembered with their respective colors. These data were recorded, and this completed the experimental session.

Results

The summary of the analysis of variance performed on the data of Exp. I is presented in Table 1. Measures used in the analysis were the total number of exposures required to reach thresholds for all four syllables of each type. None of the main effects was significant beyond the Punishment × Choice × Syllables interaction so the means for this interaction are presented in Table 2. Tukey's gap test (**10**) was used to evaluate the differences between the individual means of Table 2 at the 5% level of confidence. Those means not designated by an asterisk reflect some remaining variation among them, which is significant when tested by the *F* technique of Tukey's test ($p < .001$).

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE OF EXP. I

Source	*df*	Mean Square	*F*	*p*
Punishment	2	49.71*	1.41	
Discriminability	1	1.68*	—	
Choice	1	741.13*	21.01	.001
Syllables	1	91.13	7.15	.01
Ss within groups	132	35.27	2.77	.001
P × D	2	40.45*	1.15	
P × C	2	21.47*	—	
P × S	2	133.58	10.49	.001
D × C	1	55.12*	1.56	
D × S	1	19.01	1.49	
C × S	1	25.67	2.01	
P × D × C	2	26.53*	—	
P × D × S	2	5.56	—	
P × C × S	2	102.57	8.05	.001
D × C × S	1	8.69	—	
P × D × C × S	2	0.59	—	
Residual error	132	12.74		

* Tested by *Ss* within groups.

TABLE 2

MEAN THRESHOLDS AND RESULTS OF TUKEY'S GAP TEST OF PUNISHMENT × CHOICE × SYLLABLES INTERACTION OF EXP. I

Condition	No Forced Choice		Forced Choice	
	Shock Syllables	Nonshock Syllables	Shock Syllables	Nonshock Syllables
Punishment	18.7	17.9	19.5	12.8*
No punishment	18.2	18.0	13.9*	15.3*
Threat	18.1	17.5	13.7*	14.0*

* Separated as a straggler.

Table 2 reveals that shock syllable thresholds are significantly higher than nonshock syllable thresholds under the punishment condition when choices are forced. It should be noted that under the conditions of this experiment, not forcing choices greatly alters the character of the impending punishment in the direction of making it much less formidable.

Under the no-punishment and ambiguous-threat conditions, differences between shock and nonshock syllable thresholds are not significantly different.

Table 1 reveals that the second variable of discriminability was not effective. There were no significant differences in syllable recognition thresholds attributable to differences in discriminability.

The third variable, freedom of choice, did prove to be effective. Table 1 reveals that over-all syllable recognition thresholds are significantly lower when choices are forced than when they are not. This difference is statistically significant at the .1% level of confidence, but choice is not significant when considered apart from

its significant interaction. An inspection of Table 2 suggests that this is because of the punishment condition which appears to work at cross-purposes to it. Under no-punishment and threat conditions, forcing choices clearly results in lower thresholds.

No significant differences were found in recall of shock and nonshock syllables.

Experiment II

Method

This experiment was designed to provide some information about the fourth variable dealing with the extent of the response pool. Six additional groups of 12 *Ss* per group under each of the three punishment and two choice conditions described for Exp. I were utilized. The extent of the response pool was increased by the addition of four more syllables to those used in the first experiment: GEVAK, HIWOL, KUYEN, and LOZIM. These syllables were not present during the familiarization trials but they did appear in blue on the reference slips along with the original syllables in red and green, thereby allowing *Ss* greater latitude in their choices. Syllable differences were thus discriminably different, and choices were both forced and not forced.

These particular conditions were selected so as to simulate within the context of this experimental procedure the conditions perhaps present in McGinnies' experiment (7). It is difficult to determine whether the instructions to his *Ss* to report whatever they saw after each exposure would properly come under the forced or not forced condition of the present experimental procedure so both conditions were studied. The punishment and ambiguous-threat conditions of the present experiment can, to some extent, be considered analogous to the social-psychological threat present in McGinnies' study.

A syllable recognition threshold was obtained on each of the 12 syllables from each of the 72 *Ss* in the manner described previously. The procedure of Exp. I was modified only to the extent demanded by the inclusion of the additional syllables.

TABLE 3

Summary of Analysis of Variance of Open Set Data of Exp. II

Source	*df*	Mean Square	*F*	*p*
Punishment	2	48.43*		
Choice	1	1410.67*	17.09	.001
Syllables	2	149.20	100.13	.001
Ss within groups	66	82.56	70.74	.001
P × C	2	82.51*		
P × S	4	103.18	69.25	.001
C × S	2	9.88	6.63	.01
P × C × S	4	169.36	113.66	.001
Residual error	132	1.49		

* Tested by *Ss* within groups.

Results

A summary of the analysis of variance for Exp. II is presented in Table 3. Choice is the only main effect which is significant beyond the Punishment × Choice × Syllables interaction. This interaction is presented in Table 4. The means for all four syllables of each type shown in Table 4 were compared by Tukey's gap test using the 5% level of confidence.

A *t* test of the difference between means of the shock and nonshock syl-

TABLE 4

Mean Thresholds and Results of Tukey's Gap Test for Punishment × Choice × Syllables Interaction of Open Set Data of Exp. II

Condition	No Forced Choice			Forced Choice		
	Shock Syllables	Nonshock Syllables	Extra Syllables	Shock Syllables	Nonshock Syllables	Extra Syllables
Punishment	22.8*	14.8	17.2*	19.1*	14.8	11.8**
No punishment	22.2*	22.8*	21.1*	14.3	15.3	14.5
Threat	20.3*	21.0*	18.9*	15.8	15.6	13.8*

* Separated as a straggler.
** Separated by a gap.

lables under the punishment condition and each of the choice conditions shown in Table 4 reveals that response suppression occurs significantly only when choices are not forced ($p < .05$). When choices are forced, the difference between means approaches significance ($p < .10$). This is something of a reversal of the finding of response suppression with forced choices in the closed set data of Exp. I. When the additional syllables are present, *Ss* are able to avoid shock syllables by choosing either the nonshock or extra syllables. Since the safe syllables are chosen more frequently, the probability is increased that they will be recognized more readily. This accounts for the generally lower thresholds of nonshock and extra syllables under the punishment condition than under conditions of threat or no punishment. Why the extra syllables should be significantly higher than the nonshock syllables when choices are not forced and significantly lower when choices are forced, is difficult to explain. It may be that with forced choices the extra syllables are clearly safest; not even having been present during familiarization trials and thus not associated with shock through generalization, they are chosen most readily. This process would yield the lower thresholds. When choices are not forced, *Ss* may choose the blank as often as is necessary to avoid shock, so extra syllable thresholds are higher.

Table 4 again shows very little difference between means under the threat and no-punishment conditions.

A *t* test of the difference between the mean of the means in Table 4 reveals that the open set means are significantly higher in over-all threshold than the closed set means ($p < .05$).

One might wonder why thresholds for the extra syllables are generally lower than those for the nonshock syllables. This would seem to refute existing evidence concerning the effects of familiarity on recognition thresholds, assuming that structural differences among syllables may be ignored for the moment. It should be remembered, however, that two sets of structurally similar syllables composed the original group of eight, whereas the extra syllables had no counterparts against which they had to be compared before being identified correctly.

Experiment III

Gilchrist, Ludeman, and Lysak (2) found that both positive and negative value lowered word recognition thresholds in comparison with neutral value. On the assumption that the moderate shock present during the familiarization trials could serve as a positive accentuating factor for certain syllables, it was anticipated that under the no-punishment condition, recognition thresholds for shock syllables would be lower than for nonshock syllables. This did not occur to a significant degree. A possible explanation was that the shock syllables had not been accentuated sufficiently in the familiarization trials. This could be done by increasing the number of such trials, perhaps increasing the intensity of shock up to a point, or by reducing the ratio of shock to nonshock syllables. The latter method was used as it was considered to be the most expeditious.

A check on the recognition threshold of each syllable from the data of Exp. I indicated that there were significant differences between syllables. Therefore, for the present experiment, syllable lists were arranged so that comparisons between shock and nonshock syllables were made only between syllables which were equivalent with regard to recognition thresholds and which were similar in structure.

The data obtained from the remaining syllables were ignored.

Method

Four syllable lists were constructed and affixed to the memory drum cylinder as previously described. Each list contained the original eight syllables twice and arranged in a random order. This time, however, only two of the eight syllables appeared in a color which contrasted with the other six. Colors were again red and green.

Thresholds were obtained under a no-punishment, forced-choice condition. Reference lists with no discriminable differences between syllables were used since Exp. I indicated that syllable discriminability was not a significant variable. The original no-punishment instructions of Exp. I were given to the 24 *Ss* recruited for the experiment. The experimental procedure in all other respects was the same as that already described.

Results

Recognition threshold totals for the two shock syllables and two nonshock syllables were obtained for each *S*. The mean threshold for the shock syllables was 7.17 and the mean for the nonshock syllables was 7.33. This difference, although in the anticipated direction is not significant when tested by the *t* test for correlated measures.

An inspection of the data of this experiment revealed that this trend was almost completely neutralized by the relatively extreme scores of five *Ss* who showed the opposite result. Furthermore, since the criterion for recognition was a correct report, irrespective of whether the syllable was actually perceived or just chosen as a guess, forced-choice factors undoubtedly played a large part in the result. Therefore, Exp. IV was undertaken.

Experiment IV

Method

For this experiment the conditions and procedure of Exp. III were duplicated except that choices were not forced.

Results

The mean recognition threshold for the shock syllables was 8.67, and the mean for the nonshock syllables was 10.21. This difference, when tested by the *t* test for correlated measures, is significant ($p < .02$). When choices are not forced and when shock syllables are effectively accentuated, sensitization would seem to occur.

Discussion

To summarize briefly, these data have shown that under certain experimental conditions, punishment in the form of electric shock can motivate both in the direction of raising visual recognition thresholds and in lowering them.

The former situation provides some additional evidence in support of the response suppression hypothesis over that of perceptual defense. The *Ss* in Exp. I and II under the punishment condition were provided with a clear-cut situation where the consequences of their choices were well defined. That they avoided shock syllables and chose nonshock syllables is not surprising. To have acted otherwise would have been, at best, unrealistic. It is likely that something of the sort went on in McGinnies' experiment as well. That his was a punishing situation in the psychological sense—the forcing of naive *Ss* to report words like *bitch* and *whore*—and not merely a threatening one is suggested by the failure to obtain the same results when impending shock is merely a faint threat and not an actuality.

Some additional evidence, clearly interpretable in this same vein, has recently been provided by Lacy, Lewinger, and Adamson in their study of the effects of foreknowledge on perceptual defense (4). They find that the more incomplete the *Ss*' expectations with regard to the nature of the words they are to view, the greater the discrepancy between taboo and neutral words. It would appear that when *Ss* are told what to expect, they have less reason to suppress responses. The psy-

chological climate is structured for them and the reporting of taboo words is made much less painful. These authors point out that their data pose problems for the Howes and Solomon theory (4, p. 173) in that the theory must now account for lower thresholds for taboo words with warning, higher thresholds for taboo words when seen without warning and habituation effects, i.e., the absence of the perceptual defense phenomenon with successive words. This constitutes a problem for the Howes and Solomon theory only when the frequency explanation is cited, but much more cogent to the argument is Howes and Solomon's interpretation of McGinnies' results in terms of response suppression. This latter concept adequately accounts for the data which Lacy *et al.* present.

Postman, Bronson, and Gropper (9) also report data which indicate that relative thresholds for neutral and emotional words vary significantly with the preliminary instructions given to *S*s. An uninformed group had higher relative thresholds for taboo words than any of the groups forewarned to expect taboo words. These authors conclude that "set and selective verbal report contribute to the observed threshold differences." Response suppression is a form of selective verbal report.

When both aspects of the threshold change shown in the present study are considered together, i.e., the raised thresholds for shock syllables under punishment and the lowered thresholds when there is no punishment, an alternative way of accounting for the subception effect is also possible. Bricker and Chapanis (1) have shown that even incorrectly perceived syllables convey some information to *S*s on the basis of which they may be able to establish from which class of stimuli a certain stimulus comes, even if an incorrect verbal response is made. They suggest that common recognizable elements among the syllables may account for the GSR effects Lazarus and McCleary (5) obtained. With regard to thresholds, the sensitization effect shown in the present data is probably explainable on some such basis. If we assume that sensitization to shock syllables occurred in the Lazarus and McCleary experiment, at which time high GSR's were registered, the fact that there were no significant differences in final thresholds leads to the further assumption that some sort of response suppression occurred as well in response to the anxiety engendered by the early recognition of the shock syllables.

Summary

A series of experiments was designed to test the effects of punishment, item discriminability, freedom of choice, and extent of the response pool on syllable recognition thresholds.

In the face of impending punishment, thresholds for shock syllables were significantly higher than for nonshock syllables (*a*) when *S*s responded from a closed set of possible responses with forced choices, and (*b*) when *S*s responded from an open set with free choices. Under a no-punishment condition, when syllables were equated for recognizability and chance effects were minimized, there was some evidence to the effect that sensitization to previously shocked syllables occurs.

Under conditions of ambiguous threat, essentially equal thresholds between shock and nonshock syllables resulted.

Forcing choices generally resulted in lower over-all thresholds for both shock and nonshock syllables, whereas increasing the extent of the response pool resulted in higher over-all thresholds.

The degree of discriminability of items with respect to their affective character had no significant effect on the syllable thresholds.

The raised thresholds for shock syllables were interpreted as additional evidence favoring the response suppression hypothesis over that of perceptual defense. This process in combination with sensitization was offered as a possible alternative explanation of the subception phenomenon.

References

1. Bricker, P. D., & Chapanis, A. Do incorrectly perceived tachistoscopic stimuli convey some information? *Psychol. Rev.*, 1953, **60**, 181–188.
2. Gilchrist, J. C., Ludeman, J. F., & Lysak, W. Value as a determinant of word recognition thresholds. *J. abnorm. soc. Psychol.*, in press.
3. Howes, D. H., & Solomon, R. L. A note on McGinnies' "Emotionality and percep-

tual defense." *Psychol. Rev.*, 1950, **57**, 229–234.

4. Lacy, O. W., Lewinger, N., & Adamson, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–174.
5. Lazarus, R. S., & McCleary, R. A. Autonomic discrimination without awareness: a study of subception. *Psychol. Rev.*, 1951, **58**, 113–122.
6. McCleary, R. A., & Lazarus, R. S. Autonomic discrimination without awareness: an interim report. *J. Pers.*, 1949, **18**, 171–179.
7. McGinnies, E. Emotionality and perceptual defense. *Psychol Rev.*, 1949, **56**, 244–251.
8. McGinnies, E. Discussion of Howes and Solomon's note on "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 235–240.
9. Postman, L., Bronson, W. C., & Gropper, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1952, **48**, 215–224.
10. Tukey, J. W. Comparing individual means in the analysis of variance. *Biometrics*, 1949, **5**, 99–114.
11. Whittaker, E. M., Gilchrist, J. C., & Fischer, J. W. Perceptual defense or response suppression? *J. abnorm. soc. Psychol.*, 1952, **47**, 732–733.

(Received August 20, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

A TEST FOR CONDITIONED INHIBITION IN MOTOR LEARNING

JOHN A. STARKWEATHER AND CARL P. DUNCAN

Northwestern University

The Hullian concepts, reactive inhibition and conditioned inhibition, are now used rather widely as intervening variables by investigators of motor learning (**2, 4, 7, 12**). Reactive inhibition (I_R) is likened to a negative drive state. It accumulates during work, depresses performance, and dissipates spontaneously during rest. It is assumed to underlie in part the superiority of distributed practice over massed practice, and the gain over a long rest following massed practice (reminiscence).

Conditioned inhibition ($_sI_R$) is defined as a habit of resting for which the drive is I_R and the reinforcement is rest. These are the conditions specified by Hull (**5**). Conditioned inhibition, like I_R, depresses performance, but as a habit it does not dissipate spontaneously over rest, particularly during rests as short as the 10-min. period often assumed to permit complete or nearly complete dissipation of I_R (**3**). As a habit, $_sI_R$ is presumably subject to the laws applicable to all habits.

The several excitatory and inhibitory variables are assumed to interact as follows. Practice builds up reaction potential ($_sE_R$). If I_R and $_sI_R$ are also developed during practice, their sum subtracts from $_sE_R$ to depress performance. Performance is called effective reaction potential ($_s\bar{E}_R$); thus we may write:

$$_s\bar{E}_R = {}_sE_R - (I_R + {}_sI_R).$$

If I_R is dissipated over a rest, there would still be some depression in performance resulting from $_sI_R$ alone.

In Ammons' (**2**) application of this to motor learning, $_sI_R$ (called permanent work decrement by him) is defined as the difference between the initial postrest performance of two groups: a group that prior to the rest had worked under massed practice (MP) and a group that had worked under highly distributed practice (DP). The failure of the MP group to reminisce to the level of the DP group in initial postrest performance is assumed to be due to $_sI_R$, and this observation defines $_sI_R$. Initial postrest performance must be corrected, in both groups, for a decrement due to the need to warm up after rest.

The criterion that MP groups do not reminisce sufficiently after rest to reach the performance level of DP groups has been accepted by a number of investigators (**2, 4, 6, 7, 8, 9, 10, 12**) as indicating the presence of $_sI_R$. However, in a discussion of this matter with Dr. B. J. Underwood it was suggested that the observation defining $_sI_R$ depends upon the crucial assumption that during prerest practice the same amount of reaction potential is developed by the same amount of practice, regardless of whether that practice is massed or distributed. This assumption was explicitly stated by one of the authors (**9**), who has reported finding $_sI_R$, and it seems correct to say that the assumption was an implicit basis of the argument in all but one (**12**) of the studies reporting $_sI_R$ in motor learning. In all investigations, MP and DP groups that were to be compared after rest were given the same amount of prerest

practice. Then, if initial postrest performance of an MP group, after correction for warm-up decrement, was poorer than that of the corresponding DP group, ${}_sI_R$ was inferred. The possibility was not considered that the postrest difference in performance might have resulted from less ${}_sE_R$ being developed in the MP group during prerest practice. Wasserman (12) also gave equal prerest practice to all groups, but attempted to correct for the fact that the DP had had more reinforcement before rest. By his correction the amount of ${}_sI_R$ was reduced but not eliminated.

It is not the intention of the present paper to suggest that what has been called ${}_sE_R - {}_sI_R$ in motor learning is merely less ${}_sE_R$, although this alternative, and other criticisms (11), should be dealt with before ${}_sI_R$ is accepted. Rather, we shall show that the assumption of ${}_sI_R$ leads to a prediction that is not supported by the data to be presented. The argument is as follows.

If the same amount of prerest practice develops equal ${}_sE_R$, which in the MP group is inhibited by ${}_sI_R$, then it should be possible to compensate for the ${}_sI_R$ by giving *more* prerest practice to, and thereby developing more ${}_sE_R$ in, the MP group. If the greater amount of ${}_sE_R$ compensates exactly for the ${}_sI_R$, after a rest to permit dissipation of I_R the performance of the MP group should equal that of the DP group. In symbols, we should have (deliberately produced) after rest:

$$ {}_s\bar{E}_R \text{ (MP)} = {}_s\bar{E}_R \text{ (DP)}. $$

But according to conditioned inhibition theory,

$$ {}_s\bar{E}_R \text{ (MP)} = {}_sE_R - {}_sI_R. $$

From this it follows that immediately after rest

$$ {}_sE_R \text{ (MP)} > {}_sE_R \text{ (DP)}. $$

Now since it is assumed that the growth curve of ${}_sE_R$ is negatively accelerated (5), it follows from the above inequality of ${}_sE_R$'s that increments of ${}_sE_R$ (MP) will be *less* than increments of ${}_sE_R$ (DP) for trials beyond the point of equal performance. This leads to the prediction that if both groups practice beyond the point of equal performance under the same degree of distribution, the DP group will eventually surpass the MP group. This is the hypothesis to be tested in the experiment.

Method

Conditions.—The procedure was such that by giving a greater amount of prerest practice to an MP than to a DP group, performance on the first postrest trial was the same for both groups. This was achieved in the following way. The MP *S*s were run first. They were first given 15 distributed practice trials; throughout the experiment a distributed trial is 5-sec. work and 25-sec. rest. This initial period of distributed practice was followed immediately (after the last 25-sec. rest) by 5 min. of continuous practice. Time on target was recorded every 30 sec. during this period of massed practice. This completed the work (a total of 375 sec. of practice) for the first day. Twenty-four hours later the MP *S*s returned and were given 30 distributed trials. The 24-hr. rest should be sufficiently long to permit complete dissipation of all I_R developed during the 5 min. of massed practice.

The number of prerest trials to be given the DP group was determined by extrapolating, in the MP group data, the curve from the initial period of distributed practice to the level of performance on the first postrest trial. It was determined by this method to give the DP *S*s 30 prerest trials, each trial 5-sec. work and 25-sec. rest, for a total of 150 sec. of practice. This was done, following which the DP *S*s were given a 24-hr. rest and returned for an additional 30 distributed trials on the second day. The reason for choosing a 5–25 work-rest cycle for distributed trials was to prevent the accumulation of much I_R, and thereby, presumably, to prevent the development of ${}_sI_R$, in the DP group during prerest practice. If the work-rest cycle was correctly chosen, there should be no reminiscence in the DP group after the 24-hr. rest; the results will show that this was the case.

In addition to its use as a basis for extrapolation, the initial 15 trials of distributed practice given the MP group served two other purposes.

It permitted matching of the two groups over the first 15 trials, and it gave the MP *S*s experience with distributed practice. Thus, when the MP *S*s returned on the second day for postrest distributed practice, they did not have to practice with a work-rest cycle new to them.

Subjects.—Each group consisted of 30 male undergraduates enrolled in elementary psychology courses at Northwestern University. Students in these courses are rewarded with one point of examination credit for each day's service in experiments. None of the *S*s had had previous experience with the pursuit rotor.

Apparatus.—A 60-rpm pursuit rotor provided the task. The turntable had a radius of 14.3 cm. The target was a brass disk, .87 cm. in radius, set flush with the surface of the table; the center of the disk was 7.8 cm. from the center of the table. The *S* pursued the target with a hinged brass stylus; the hinge prevented *S* from staying on the target by pressing down on the stylus. When the stylus was in contact with the target a circuit was completed to two .01-sec. clocks. The clocks alternated in recording every 30 sec. to permit *E*'s recording of time-on-target scores for 30-sec. trials during continuous practice. During distributed practice the rotor was automatically timed to run for 7 sec. and stop for 23 sec. The clocks recorded only during the middle 5 sec. of the 7-sec. rotation; thus *S* could not score during the 1-sec. periods of acceleration or deceleration.

Instructions.—All *S*s were given general instructions on the nature of the task, including suggestions and a demonstration of the best way to practice. The 5–25 work-rest cycle was demonstrated, and they were told there would be further instructions between trials later on. After the initial 15 trials of distributed practice, the MP *S*s were told, during the last 25-sec. rest, that the rotor would run continuously for several minutes. When they came in on the second day, all *S*s were told that practice on this day would be distributed.

Additional groups.—After the data for the original two groups were obtained, a partial check upon the results was made by running two additional groups of 15 *S*s each. The groups were treated identically to the first two groups except that during the postrest session (the second day) they were given 7 min. of continuous practice.

Notation.—Since the first 15 prerest trials of distributed practice were the same for all groups, they are ignored in the notation. Each group is designated by the kind of practice immediately before and after rest. Thus Group MP–DP is that group given continuous practice immedi-

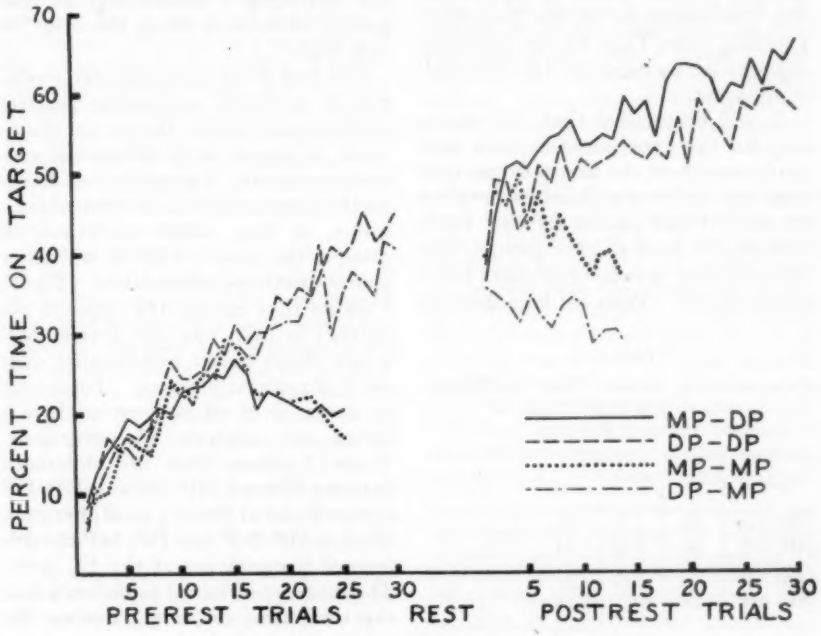


FIG. 1. Performance on all trials for all groups

ately before, and distributed practice after, rest. In addition, Groups MP–DP and DP–DP, each containing 30 *Ss*, will be called the first pair of groups, while those run subsequently (MP–MP and DP–MP, 15 *Ss* each) will be referred to as the second pair.

Results

Figure 1 shows the performance curves for all groups on all trials. In the figure, scores are expressed in percentages, thus permitting comparison of MP (30 sec.) with DP (5 sec.) trials. It may be seen that all groups appear to be well matched during the first 15 prerest distributed practice trials. This was supported by statistical analysis; there was no significant difference, on total score for these 15 trials, either between Groups MP–DP and DP–DP ($t = .42$), or between Groups MP–MP and DP–MP ($t = .02$). Tests for homogeneity of variance resulted in *F*'s of 2.05 for the first pair of groups, and 1.05 for the second pair of groups; neither value is significant. For the remainder of the first day's practice, after Trial 15, the typically superior performance of DP over MP is apparent.

It will be recalled that it is necessary for this experiment to show that performance on the first postrest trial does not differ significantly between prerest MP and prerest DP *Ss*. Table 1 shows, for both pairs of groups, that this provision was met; neither *t* value is significant. Tests for homogeneity of variance gave *F*'s of 1.18 and 1.13 for the first and second pairs of groups respectively, neither significant. Belief that the comparable performance levels of Groups MP–DP and DP–DP on the first postrest trial were not merely an accident is strengthened by the replication afforded by the subsequently run Groups MP–MP and DP–MP; Fig. 1 shows that all four groups performed about the same on this trial.

TABLE 1

Comparison of Means (Time on Target) during the First Trial of Postrest Practice

Group	*N*	Mean (sec.)	σ_M	*t*	*p*
MP–DP	30	1.81	.13	.554	.50
DP–DP	30	1.92	.15		
MP–MP	15	11.89	1.21	.673	.50
DP–MP	15	10.77	1.14		

TABLE 2

Analysis of Covariance of Total Time on Target Over All Postrest Trials

Source	*df*	Adjusted Mean Square	*F*
Group MP–DP versus DP–DP			
Between groups	1	860.17	4.50*
Within groups	57	191.30	
Group MP–MP versus DP–MP			
Between groups	1	13406.19	8.88**
Within groups	27	1509.09	

* Significant at the 5% level.
** Significant at the 1% level.

The test of the experimental prediction is made by comparing postrest performance under the same conditions of groups with differential prerest conditions. The prediction, based on the assumption of conditioned inhibition, is that, within each pair of groups, the prerest MP *Ss* will show *poorer* postrest performance. Figure 1 shows that exactly the opposite occurred; in each pair the prerest MP group shows better performance over most of postrest practice. Total time on target over all postrest trials was tested by analysis of covariance. Table 2 shows that the difference between Groups MP–DP and DP–DP is significant at the 5% level; between Groups MP–MP and DP–MP the difference is significant at the 1% level. The measure of initial proficiency was the total time on target during the

first 15 trials of prerest practice. Tests for homogeneity of variance gave F's of 1.14 and 1.20 for the first and second pairs of groups respectively, neither significant. It is clear that under either condition of postrest practice the prerest MP *S*s show superior performance.

Discussion

Perhaps the objection most likely to be raised to the experiment will be that the matching of performance on the first postrest trial of prerest MP and DP groups is spurious because no correction was made for warm-up or loss of set in the prerest MP *S*s. It has been assumed by some (1, 2) that during a long rest there is a loss of set in prerest MP *S*s. On this assumption, performance on the first postrest trial, even though it may represent a considerable gain over the last prerest trial, underestimates the total reminiscence gain by a prerest MP group, because performance on that trial is depressed by the need to warm up. The presence of a warm-up decrement is presumably indicated by a steep slope in the postrest curve for the first few (usually two to four) trials, with a subsequently more gradual rise as practice continues. It has therefore been argued that the true initial level of postrest performance in a prerest MP group is the highest point of the relatively steep portion of the curve (2).

If these assumptions concerning warm-up were known to be correct, it could be argued that the true initial level of postrest performance has been underestimated for the prerest MP groups in the present experiment. Then the superior performance over most of the postrest period by prerest MP *S*s would be no problem for the defender of conditioned inhibition. He could even argue that the performance superiority would be greater than that found, were it not for the presence of ${}_{S}I_{R}$.[1] However, objecting to the present experiment on the basis of failure to correct for warm-up decrement requires more evidence for the assumptions made about the warm-up process than exists at present. The results on warm-up have been quite inconsistent. In a recent paper (1) devoted specifically to studying warm-up, there was as much warm-up in a DP group as in an MP group after the first 24-hr. rest. At least for the present, our contention that the results provide a difficulty for conditioned inhibition theory cannot be negated by an appeal to warm-up alone.

The question might also be raised as to whether ${}_{S}I_{R}$ extinguishes very rapidly, which, if true, would account for the postrest superiority of prerest MP groups. For this to be a valid objection, ${}_{S}I_{R}$ would have to extinguish in the first few postrest trials to account for the results, and in the groups practicing under MP after rest, would have to extinguish under the same conditions under which it is assumed to develop. Actually, the whole problem of extinction of ${}_{S}I_{R}$ is somewhat paradoxical in that omission of reinforcement (rest) would presumably be achieved only under continuous practice in which *S* would be prevented from taking even a "bootleg" rest, i.e., a lessening of effort while still practicing, which has been assumed (8) to be a reinforcement for ${}_{S}I_{R}$. This does not seem feasible.

We should like now to re-emphasize a point made in the introduction to this paper. It was pointed out that the operations used to define ${}_{S}I_{R}$ in motor learning depend upon the crucial assumption that the same amount of reaction potential is developed by the same amount of practice, regardless of whether practice is massed or distributed. It would actually be more parsimonious to assume that less ${}_{S}E_{R}$ is developed in the presence of I_R, more parsimonious in that one less concept (${}_{S}I_{R}$) would be needed to ac-

[1] Such an argument would, however, be contradictory. If prerest MP *S*s have developed a greater amount of reaction potential, then, again because of the assumption that the development of reaction potential is negatively accelerated, the postrest performance curve of a prerest DP group should have a steeper slope than the curve for the corresponding prerest MP group. Figure 1 shows no evidence for this.

count for the failure of an MP group to reminisce to the level of a DP group after rest. It should be pointed out, however, that the results of the present experiment do not support an assumption that I_R slows the rate of development of ${}_sE_R$. To claim this we would also have to account for the lack of a difference on the first postrest trial. For if the matching on the first postrest trial indicates equal ${}_sE_R$ with no ${}_sI_R$, groups practicing under the same conditions after rest should not differ in performance. This was not the case.

The present paper is not intended to be a criticism of the concept of reactive inhibition. I_R is a fruitful concept but it would be even more so if more were known about its characteristics. In particular, the possibility that I_R decreases late in practice, as indicated by Kimble and Shatel (10), should be noted. If, arguing from their findings, it can be assumed that susceptibility to I_R is reduced even after a single experience with I_R, the differences between the postrest performance levels of each of our two pairs of groups are explained. Both prerest MP groups did better in postrest practice than the corresponding prerest DP groups, because the MP *S*s were not as susceptible to the I_R developed by postrest practice. And this effect was more striking in the case of the two groups working under continuous practice after rest, where a large amount of I_R developed rapidly.

Finally, the present paper is in no sense a criticism of the concept of ${}_sI_R$ as it is used in conditioning. But, as has been pointed out elsewhere (11), in motor learning the concept is subject to several criticisms; no data yet presented provide clear-cut evidence for ${}_sI_R$ in motor learning.

Summary

The concept of conditioned inhibition as it is used in motor learning was criticized on both theoretical and experimental grounds. It was shown that conditioned inhibition theory predicts that if one group is given massed practice and another group given distributed practice before a long rest, and after the rest both groups practice under the same degree of distribution, late in postrest practice the prerest distributed group should be higher, as long as both groups had the same initial level of postrest performance. In an experiment designed to test this prediction, matching of initial levels of postrest performance was achieved by giving more practice before rest to the massed than to the distributed group. The results contradicted the prediction; except for the first few trials, the prerest massed group significantly exceeded the prerest distributed group in postrest performance. It was suggested that conditioned inhibition has not yet been demonstrated in motor learning.

References

1. Adams, J. A. Warm-up decrement in performance on the pursuit-rotor. *Amer. J. Psychol.*, 1952, **65**, 404–414.
2. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
3. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
4. Barch, A. A. Work decrements in the performance of a pursuit task arising from relatively short periods of massed practice. *Iowa Acad. Sci. Proc.*, 1951, **58**, 313–320.
5. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
6. Hull, C. L. *Essentials of behavior.* New Haven: Yale Univ. Press, 1951.
7. Kimble, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.
8. Kimble, G. A. Performance and reminiscence in motor learning as a function of the degree of distribution of practice. *J. exp. Psychol.*, 1949, **39**, 500–510.
9. Kimble, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, **40**, 248–253.
10. Kimble, G. A., & Shatel, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
11. Underwood, B. J. Learning. *Annu. Rev. Psychol.*, 1953, **4**, 31–58.
12. Wasserman, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, **42**, 162–172.

(Received August 13, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

WARM-UP IN MASSED AND DISTRIBUTED PURSUIT ROTOR PERFORMANCE[1]

ABRAM M. BARCH

State University of Iowa[2]

Ammons (2) in a theoretical formulation of warm-up in motor learning predicted that the warm-up decrements shown by a massed practice group would first increase to a maximum and then decrease as the number of practice sessions was increased. Adams (1) tested this prediction in an experiment concerned with warm-up under both massed and well-distributed practice conditions. The massed group displayed no consistent trend of warm-up decrements.

Because the extent of the distribution of practice resulted in absolute warm-up losses in performance level, Adams was able to utilize two measures of warm-up decrement with his distributed group: (*a*) decrement method—the difference between the performance level on the last prerest trial and the first postrest trial; (*b*) initial increase method—the difference between the performance level at the peak of the sharp initial rise of the postrest performance curve and the level of the first postrest trial. Both measures indicated that the amount of warm-up decrement under well-distributed practice conditions decreased as the number of practice sessions increased. In addition, these two different operational definitions agreed quite closely in absolute magnitude as well as in trend.

The failure of the massed practice group to display a consistent trend of warm-up loss plus the consistent trend for the distributed group suggested that a reconsideration of the warm-up situation might be in order.

Neglecting the procedure used by Ammons for measuring warm-up, only two assumptions are necessary for his prediction. Assumption 1 (2, pp. 271 f.) of his system suggests that warm-up decrement will increase with practice if warm-up decrement can be regarded as a loss in "set" (advantageous postural adjustments) during the rest period. "The more practice before rest, the more [*of these adjustments*] there will be to lose." Assumption 4 (2, p. 274) indicates that, "The more times practice is resumed, the easier it will be to warm-up; *i.e.*, there will be learning of 'methods of warm-up.'"

These assumptions can apply to distributed practice as readily as they can apply to massed practice. The ensuing prediction that warm-up decrements in a well-distributed practice group would first increase to a maximum and then decrease as the number of practice sessions increases appears to be contrary to the results obtained by Adams. He found a continuously declining function. However, his distributed group had come near to asymptotic performance before a rest was introduced in order to measure warm-up decrement. Therefore, his data might be regarded as comprising the decreasing portion of the pre-

[1] This article is based on a thesis submitted to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the M.A. degree. The writer wishes to express his gratitude to Dr. Don Lewis for aid and assistance throughout the course of this investigation.

[2] Now at the Human Resources Research Office, George Washington University.

dicted curve. To check this point, frequent measurements of warm-up decrements during the early stages of acquisition of the motor skill are necessary.

No attempt is made here to predict warm-up under massed practice conditions since it is felt that well-distributed practice, not massed practice, represents the simpler case. This point will be considered more fully in the discussion section.

The purpose of this investigation, then, was to study: (*a*) the change in the warm-up loss displayed by a well-distributed practice group and a massed practice group on a rotary pursuit task as the number of practice sessions was increased, and (*b*) the relation between the initial increase method and the decrement method of measuring the amount of warm-up.

The expectation was: As the number of practice sessions increased, the warm-up loss displayed by the distributed group would increase to a maximum and then decrease.

Method

Subjects.—The *S*s were 52 college males. They were assigned at random to distributed practice (Group D) or to massed practice (Group M). The data of three *S*s could not be used in the results because of apparatus difficulty or failure to serve on the last day of the experiment. There were 25 *S*s in Group M and 24 *S*s in Group D.

Apparatus and task.—In order to increase the difficulty of the task and thus slow down the learning, the Epicyclic Pursuit Rotor with an epicyclic target path was used rather than the more common Koerth Pursuit-Rotor. This pursuit apparatus consists of a large Bakelite disk, 14 in. in diameter, and a smaller satellite disk, 5½ in. in diameter. The satellite disk is mounted flush with the large disk and toward the outer rim. The outer edge of the satellite disk is ¾ in. from the outer edge of the large disk. The small disk rotates about a center 3½ in. from the center of the large disk. The brass target, ¾ in. in diameter, is mounted flush with the small disk. The outer edge of the target is ½ in. from the outer edge of the small disk.

Both disks rotate at 40 rpm. During each clockwise rotation of the large disk, the satellite revolves once in the same direction. Thus, the path traced by the target is an epicyclic or somewhat heart-shaped pattern with the loop of the pattern at the point nearest *S*. The task of *S* was to keep a hinged stylus in contact with the target.

Design and procedure.—Both groups practiced 30 trials a period for two periods in blocks of six trials each with a 5-min. rest period between each block. Group D returned for a third practice period of only six trials. Forty-eight hours intervened between the first and second practice period and 24 hr. elapsed between the second and third period. Trials were 15 sec. in length. Intertrial rests were 45 sec. for Group D and 0 sec. for Group M. (A 5-min. rest period was chosen after inspection of Ammons' [4] curve of warm-up decrement as a function of interpolated rest. Relatively little difference appeared between a 5-min. rest period and the longer rest periods.)

Two *S*s served at the same time. The *S*s in Group D alternated after every trial; *S*s in Group M alternated after each block of trials. The *S*s in both groups sat in a rest booth just outside the experimental booth and read magazines during the rest periods. During a block of trials the nonperforming Group D *S*s stood just outside the experimental booth. No smoking was allowed and *S*s could not observe performance of the partners. This procedure of requiring *S*s to leave the experimental booth as frequently as possible was designed to force them to make postural adjustments.

A red warning light was presented 3 sec. before each trial or block of trials. The *S* then picked up the stylus, placed it on the target, and prepared to commence tracking.

The pattern of movement of the target and the manner in which the stylus was to be held were demonstrated to *S*s. The *S*s were told that their score would be the total amount of time that their stylus was in contact with the target during a trial period.

Recording and timing.—Time on target was recorded for the last 14 sec. of each 15-sec. trial. This procedure duplicates Adams' (1) use of a free revolution. Two clocks were used to record the scores of Group M. Readings for both groups were taken to the nearest .01 sec. Presentation of the warning light, trials, and intertrial rests was completely automatic. The 5-min. rest periods were timed by stop watch.

Results

Figure 1 shows the amount of warm-up for Group D after the various rest

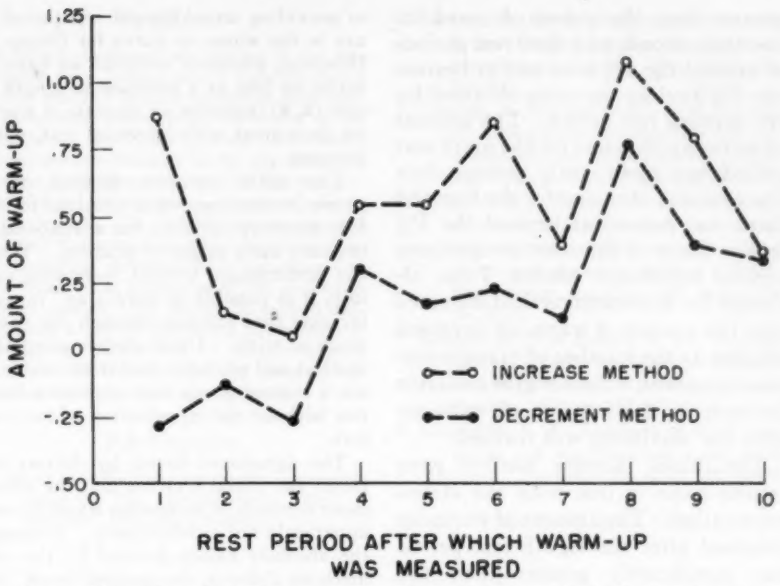


FIG. 1. Comparison of the decrement and the initial increase methods of measuring warm-up losses of Group D

periods as measured by the initial increase method and the decrement method. Except for the first point of the curve, the two methods gave similar results. The value obtained after the first rest period by using the initial increase method was probably affected by the rapid increase in proficiency at that point and did not truly represent a recovery or relearning process.

The general trend of the two curves in Fig. 1 was an increase to a maximum and then a decrease; however, a statistical test was needed for a more rigorous determination.

A simple analysis of variance of repeated measures (6) was used to test the hypothesis that the mean amount of warm-up associated with each rest period throughout the experiment was the same for all the rest periods. This hypothesis was rejected at beyond the 5% level of confidence for the decrement method ($F = 2.06$ for 9 and 207 *df*) and at beyond the 1% level for the initial increase method ($F = 2.78$ for 9 and 207 *df*). Table 1 provides a summary of the analysis for the initial increase method.

With the decrement method, the amount of warm-up obtained for the eighth rest period was significantly

TABLE 1

ANALYSIS OF VARIANCE OF AMOUNT OF WARM-UP FOR GROUP D AFTER THE VARIOUS REST PERIODS (INITIAL INCREASE METHOD)

Source	*df*	Mean Sum of Squares	*F*
Warm-up occasions	9	2.99	2.78**
Between *Ss*	23	1.08	
Remainder	207	1.07	
Total	239		

** Significant at the 1% level.

greater than the values obtained for the first, second, and third rest periods at beyond the 1% level and at beyond the 5% level for the value obtained for the seventh rest period. The amount of warm-up obtained for the ninth rest period was significantly greater than the amounts obtained for the first and third rest periods at beyond the 5% level. None of the other comparisons yielded significant results. Thus, although the decrement method indicated that the amount of warm-up increased reliably as the number of practice sessions increased, it did not give a reliable decrease in the amount of warm-up after the maximum was reached.

The initial increase method gave results more in line with the stated expectation. The amount of warm-up obtained after the eighth rest period was significantly greater than the amounts obtained after the second, third, seventh, and tenth rest periods at the 1% level and after the fifth rest period at the 5% level. There was evidence, then, for a reliable increase to a peak and a reliable decrease after that peak.

The hypothesis that the warm-up decrements for Group M were the same after all rest periods could not be rejected at the 5% level ($F = 1.11$ for 8 and 192 *df*).

Discussion

The expectation that the amount of warm-up in Group D would increase to a maximum and then decrease as the number of practice sessions was increased received considerable support. More practice was probably necessary to yield a reliable decrease, after the maximum, with the decrement method.

It was unfortunate that the only rest period indicating a significant decrease from the point of maximum warm-up loss, the tenth rest period, was 24 hr. in duration rather than 5 min. This rest period was added merely for the purpose of providing an additional rest period to use in the warm-up curve for Group D. However, whatever evidence we have on warm-up loss as a function of length of rest (3, 4) indicates an *increase* in warm-up decrement with increased rest, not a decrease.

The initial increase method, as is shown by the large value obtained for the first warm-up occasion, has a weakness in the very early stages of practice. When the performance level is increasing rapidly, it is possible to obtain an "initial" increase that extends through the entire block of trials. Later studies using this method will probably find it advisable to use a control group that continues practice without the introduction of the long rest.

The agreement found by Adams between the initial increase and the decrement methods of measuring warm-up was apparently not a coincidence. Although the absolute values derived by the two methods differed, the general trend (except for the first point) was the same.

No reliable trend of warm-up loss was predicted for Group M and none was found. This added some weight to the claim made in the introduction that well-distributed practice, not massed practice, represents the simpler case for the study of the warm-up phenomenon. Massed practice, meaning practice without inter-trial rests, appears to introduce several complicating factors. For one thing, there is a marked difference between the last prerest trial of a massed practice group and the early postrest trials in terms of the stimulation (both kinesthetic and motivational) furnished by *S*. During the last prerest trial *S* was performing with an accumulation of inhibitory potential; during the early postrest trials he is not. In addition, reminiscence may interact with the warm-up process in some way.

Readers desiring additional experimental support for the present results may apply the two methods of measuring warm-up to the performance curves for massed and distributed pursuit rotor performance in a recent study by Kimble and Shatel (5), which appeared after the present study was completed.

Summary

Warm-up decrement was studied in the performance of a massed and a well-distributed practice group in an epicyclic pursuit rotor task. Warm-up was measured by the extent of the sharp initial rise in the postrest performance (initial increase method) or by the difference between the performance level on the last prerest trial and on the first postrest trial (decrement method). The initial increase method indicated that warm-up under distributed practice conditions increased to a maximum and then decreased as a function of the number of practice sessions. The decrement method indicated only a reliable increase in the amount of warm-up as the number of practice sessions increased. No reliable trend for warm-up was obtained for the massed practice group.

References

1. Adams, J. Warm-up decrement in performance on the pursuit-rotor. *Amer. J. Psychol.*, 1952, **65**, 404–414.
2. Ammons, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.
3. Ammons, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393–411.
4. Barch, A. M. Warm-up in epicyclic pursuit performance under massed and distributed practice conditions as a function of the number of practice sessions. *USAF, Hum. Resour. Res. Cent., Res. Bull.*, 1952, No. 52–8.
5. Kimble, G. A., & Shatel, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355–359.
6. Lindquist, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.

(Received July 24, 1953)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

TWO PROCESSES IN PERCEPTUAL LEARNING[1]

FRANKLIN P. KILPATRICK

Princeton University

It can be stated categorically that learning plays a role in visual space perception, even though there is considerable disagreement concerning the importance of that role. The author has placed extreme emphasis on the side of learning (e.g., **8, 9, 10**) in contrast to some other views (e.g., **3, 11, 12, 13**). Perhaps, though, it would prove profitable to lay aside, at least for a time, the question of "how much?" and pay a bit more attention to the question of "how?"

Work with monocular distorted rooms designed by Adelbert Ames, Jr. (**6**) has led the author to the formulation of certain notions concerning possible learning processes involved in visual space perception. Such rooms were built originally as an empirical test of a limited case of the hypothesis of equivalent configurations, that is, for the case of a three-dimensional object monocularly observed (**7**). A monocular distorted room is nonrectangular, but built in such a way that each part of the room subtends the same visual angle at the nodal point of an eye placed at the chosen observation point as would the corresponding part of a particular rectangular room (reference room) observed from the same point. If this prescription is followed, the stimulus pattern on *O*'s retina will be the same for the distorted room, of whatever shape, and the undistorted reference room, except, of course, for very slight differences due to unavoidable defects in construction and inability to place *O*'s eye *exactly* at the prescribed observation point. Viewed by a naive *O* with one eye from the prescribed location, such distorted rooms are perceived as being like the reference room, that is, "normal" or very nearly "normal" in shape. Sloping walls are seen as vertical, sloping floors and ceilings are seen as level, and windows of different sizes and shapes are seen as being alike. Apparently no amount of time spent in just looking at such a static configuration has any effect in altering the way in which it is seen. However, and this is the important point for the purposes of this paper, *O* can *learn to see* more and more of the distorted shape of the room, even though the stimulus pattern on his retina remains unchanged.

Occurrence of this learned alteration in visual space perception was first noted in connection with the behavioral testing of *O*'s perception of a distorted room. The *O* is given a wand and asked to touch various parts of the room while keeping his eye at the observation point; he is initially unsuccessful in his efforts. His actions are those which would work in an undistorted room, and are in accord with the verbal reports of how the room is seen. Gradually, however, as he continues to act with the wand and to experience the consequences of his actions, he becomes increasingly successful in touching various parts of the room. More important, *O*s reported that concurrent with this expe-

[1] The research reported in this paper was supported by the Office of Naval Research as a portion of a larger project initiated by the Professional Division, Bureau of Medicine and Surgery. However, the opinions expressed are the author's, and should not be construed as representing the opinions or policy of the naval service.

rience, there was an alteration of the appearance of part or all of the room in the direction of its true shape. These initial findings were substantiated by several hundred observations in undergraduate laboratory classes and elsewhere.

Hoffman (4) was the first to deal with and measure the phenomenon in controlled experiments. His findings verified the earlier observational results, and showed further that the perceptual alteration was gradual and tended to increase with added practice, at least within the limits of the relatively small amounts of practice his *O*s had.[2] In this experiment, distorted rooms were used as a means of obtaining evidence concerning two theoretical formulations in the area of learning in visual space perception.

Statement of the Problem

It was briefly suggested in an earlier paper (7) that the perceptual modification that occurs in the distorted rooms involves at least two learning processes, overlapping and affecting one another, but logically and perhaps experimentally distinguishable nevertheless.

Reorganizational learning.—The first process is one in which an unconscious reweighting of visual cues occurs. Unavoidably present in every such room are cues that we are able to perceive in ways not in harmony with the dominant percept; thus they have the potential of functioning as "give-away" cues. These cues, which "give away" to some degree the distortions present in the room but are not at first utilized, are given greater and greater weight as the incompatibilities between the original percept and the consequences of the practice continue to be registered, with the result that changes in the apparent shape of the room occur. The cues referred to are those which are present because of minor, and almost impossible to correct, defects in construction, lighting, placement of *O*'s eye, etc.; also, of course, they may be deliberately introduced. Probably "reorganizational learning" is not the ideal name for this process, but it does suggest that what is learned is a new way of organizing into a whole the complex of previously established cue-percept relationships. It is the usual explanation offered for perceptual modification such as that in the distorted room, and probably few people would argue that such a process does not occur.

Formative learning.—The second process, which we shall call "formative learning," is, however, far more crucial than the first in terms of perceptual theory. It is believed to consist of an actual learned alteration in the way in which a given stimulus pattern is perceived, a new perception which is not dependent on the utilization of give-away cues. It is a learning process in which the perception of the basic configuration related to that particular distorted room, the reference room, or equivalent distorted rooms of any shape, is modified. Reorganizational learning shakes up and rearranges the marbles in the bag; formative learning puts new marbles in the bag, and, if we are correct in our view, explains how most, if not all, of the marbles got in the bag in the first place.

From this two-process theory certain deductions concerning perceptual learning in distorted rooms may be made and tested as a means of checking the adequacy of the theory. If

[2] It should be noted that, so far as the author knows, no one has practiced sufficiently long to see the full, true shape of the distorted room. Perhaps it could be done, but at present the only evidence is for varying degrees of modification of perception in the direction of the true shape.

one had two or more equivalent rooms of different shapes and *O* were given sufficient training at one of them (Room 1) to perceive to some degree its true shape, and were then transferred to an equivalent room of an entirely different shape (Room 2), one would expect the following:

a. If the only learning that took place was reorganizational, that is, dependent on give-away cues, there would be no transfer of the newly learned percept from Room 1 to Room 2, because no give-away cues to the Room 1 shape could be available in Room 2. Thus, Room 2 would be seen as normal after the Room 1 learning.

b. If, on the other hand, as the theory states, formative learning is also involved, there should be some transfer of the newly learned percept. That is, one would expect Room 2 to be seen as "normal" no longer, but to be seen at least to some degree as being shaped like Room 1.

The major purpose of this experiment was to show, if possible, that formative learning does occur in visual space perception.

A second problem in theory was included in the experiment since a rough check on its adequacy could be included without much trouble. This second problem is concerned with the role of action. The author (**8, 9**) and others (**1, 2, 5**) with a similar point of view have stated that actions and their consequences are instrumental in the "construction" of what we see. What exactly is meant by action has not been well specified, but the implication has always been, probably intentionally, that it means gross, overt bodily activity, the doing of something. This implies that one's perception of a distorted room would remain unmodified, or very nearly so, if he merely observed the room's interior while someone else carried out the prescribed action, such as feeling around the interior of the room with a wand. Of course, *O* would not be completely inactive under these conditions. There would be eye movement and minor bodily adjustments, but generally speaking his activity would be minimized to an extent that, on the basis of theory as stated, the author would be forced to predict that little or no perceptual learning would occur under such circumstances.

FIG. 1. The cubical "normal," or N room; caster-mounted chair with face piece in place against the front panel

METHOD

Apparatus

The basic apparatus consisted of three rooms designed to be monocularly equivalent, two of them distorted and one a 4 × 4-ft. cubical room (the "normal" or N room). One of the distorted rooms is larger on *O*'s left than on his right (the L room), and the other is larger at the top than at the bottom (the T room).[3] In constructing the three rooms used in this experiment great care

[3] Construction drawings are contained in (**5**).

was taken to eliminate nonequivalent visual cues such as grain in the wood, nail or screw heads, shiny reflecting surfaces, etc. The interiors were painted a flat gray with the exception of the window panels, which were painted a flat black. Plywood panels were placed over the fronts of the rooms. The rooms were arranged in a blacked out experimental room so that their fronts formed a U, the distorted rooms facing one another, and the front of the N room forming the bottom of the U.

Lighting for each room was from a number of 7-w. bulbs placed inside the room along the top of the front panel, above *O*'s head and out of his visual field. Extreme care was taken to equate the brightnesses of all parts of each room, and to equate the rooms with each other.

An armless straight-back swivel chair on casters, on which *O* sat straddling the seat and facing the back of the chair, was provided. A face piece at the top of the chair back provided a means of guiding *O*'s face to the observation position and of preventing undue head movement. In the front panel of each room an oval opening was cut that was just large enough to admit *O*'s nose and allow him to observe the interior of the room with the right eye when his nose was as far down and to the left as it would go. This use of a nose rest permitted more accurate placement of the eye than would a chin rest. Since these observation holes were at the same height

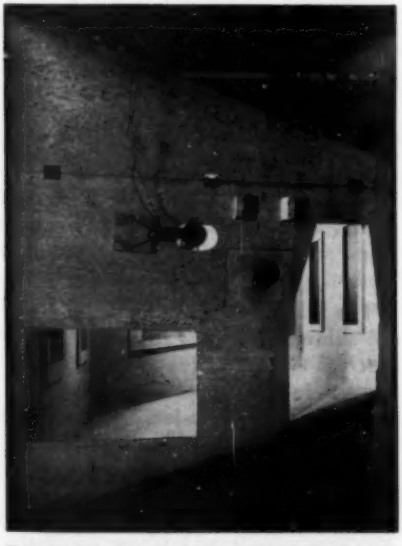


FIG. 2. The left-distorted or L room, with the spotlight in place

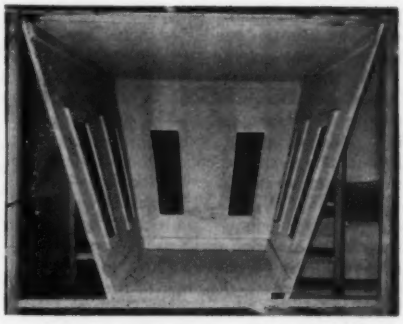


FIG. 3. The top-distorted or T room, with the front panel removed

in each of the rooms, a single adjustment of the face piece sufficed for all. Thus *O* could be moved from room to room while seated on the chair, and in each case simply putting his face back in the face piece insured the proper observational position.

The L room, which was to be the practice room, was provided with some additional equipment. A number of sponge rubber balls, covered with scotch tape so as not to mark the interior of the room when thrown against the walls, were provided. A small spotlight on a swivel was pointed into the room through a hole slightly to the left and above the observation hole. It was placed as near to the observation point as possible so that the alteration in the visual angle subtended by the spot as it was aimed at various parts of the room would be very little from *O*'s point of view. This spot of light, about 3 in. in diameter when the spotlight was aimed at a point directly opposite the hole into which it was inserted, served as a target at which to throw the balls. A light bamboo wand 4 ft. long and covered with several thicknesses of flannel at one end was also provided. In addition, most of the bottom third of the front panel was cut out to permit both *O* and *E* to have free access to the interior of the room with their hands.

A tape recorder was also used.

Observers

The *O*s were 12 young adults, 10 males and 2 females, all with normal vision without glasses.

Procedure

Equivalence testing.—In order to be sure that any distortions in appearance of the rooms following the learning sessions could be attributed to learning, it was necessary first to establish for each *O* that the L, N, and T rooms were, in fact, equivalent and all seen as "normal" under the

prescribed conditions. This was done through equivalence testing conducted immediately prior to the learning sessions for each *O*.

The *O* first was shown line drawings of the L, T, and N rooms, and their shapes were fully described. He was told that he would be seated in a chair and moved from room to room, that he would be given two looks at each of them, and that his task in each case was simply to identify the room as L, N, or T. Then a mask was placed over his left eye, he was led with his eyes closed into the blacked out experimental room, seated in the chair, and given two 10-sec. looks at each of the three rooms, the order of these six looks being randomly determined. Following the viewing and attempted identification, he was asked to describe how the rooms looked and, if he had felt he could tell them apart, the ways in which they looked different.

The results show that the rooms are seen as normal and are monocularly equivalent under the prescribed viewing conditions;[4] thus, any different results following a learning session may be treated as change in perception, provided one trusts *O*'s reports in both instances. Of the 72 room identifications, 51 were incorrect and 21 were correct. This is just slightly worse than chance. Of the 21 correct identifications, 15 were of the N room; this is not surprising in view of the fact that all 12 of the *O*s stated after viewing that the rooms all looked alike, that is, quite normal. In fact, almost half of the *O*s either "gave up" or else disbelieved *E*, three responded "normal" all six times and two others did so five times in spite of having been told in advance that they would view each of the three rooms twice. The L room was correctly identified four times, twice by the same *O*. He stated, however, that he was just guessing and that it looked normal. The T room was correctly identified only twice, and then by different *O*s.

The learning sessions.—All learning sessions were conducted at the L room, and consisted of a number of throws of the ball at the spot of light at various locations on the back wall, and of tracing the outline of the back wall four times with the wand. Half of the *O*s, the action group (Group A), tossed the ball and manipulated the wand themselves. The other half, the no-action group (Group NA), merely observed the interior of the L room while *E* or his assistant threw the ball or handled the stick. The *O* was asked to call out "hit" when he saw the ball hit the spot. This was not done for data purposes, but simply as a device to maintain *O*'s interest and motivation.

There were four learning sessions separated by rest periods of about 2 min. Immediately after each learning session, *O* was asked to observe and describe the interior of one of the three rooms: the L room after Session 1, the L room again after Session 2, the N room or the T room after Session 3, and the remaining room after Session 4. Half of the *O*s saw the T room after the third session and the N room after the fourth; for the other half the order was the reverse. All assignment to groups and orders was random.

Room descriptions were elicited in a standard way with the question, "please describe the shape of the room as you see it, that is, not how you might guess or know it to be, but how you actually see it." It was desired that every *O* describe at least the back wall, the windows in the back wall, the floor, and the ceiling, so omissions were filled in with questions in the form, "how does (the back wall) look to you?" All instructions, questions, and answers were tape recorded and transcribed verbatim for analysis.

When the *O*s were moved from the L room to one of the others, they knew they were being shifted to either the N or the T room, but not which one.

The four learning sessions differed only in the ball throwing; in Session 1 there were five throws at each of five positions of the spot, in Session 2 there were five throws at nine positions, in Session 3 there were three throws at nine positions, and Session 4 was the same as Session 3.

Postexperimental interview.—Following the final observation, *O* was asked to discuss the nature of his experiences, particularly any alterations in what he saw and any discrepancies that he felt existed between what he saw and what he otherwise judged to be so. This material also was tape recorded and transcribed.

Results

The main results of the experiment are summarized in Table 1. Of the six *O*s in Group A, five reported modification in the appearance of the L room after the first learning session. All six reported such modification for the L room after the second session. There were marked individual differences in both amount and type of change reported, but in all cases the change was in the direction of the true

[4] As far as the author knows, this is the first time that the equivalence of monocular distorted rooms has been demonstrated in a controlled experiment. The equivalence of binocular distorted rooms was experimentally verified and reported in a paper by Ittelson and Kilpatrick (7).

TABLE 1

REPORTED CHANGES IN APPEARANCE OF THE THREE ROOMS AFTER LEARNING SESSIONS FOR THE ACTION GROUP ($N = 6$) AND THE NO-ACTION GROUP ($N = 6$)

Condition of Observation	Reported Room Appearance			
	A Group		NA Group	
	No Change	Changed in One or More Aspects	No Change	Changed in One or More Aspects
L room				
After 1st learning session	1	5	2	4
After 2nd learning session		6		6
Transfer				
At the N room	1	5	2	4
At the T room		6	1	5

shape of the room. All six reported seeing to a marked degree the slope of the floor, and five of the six reported additional changes, such as the slope of the ceiling, the shape and slant of the back wall, the difference in size and shape of the two windows in the back wall, etc.

These alterations in appearance carried over to both the N and the T rooms,[5] according to *O*s' reports, for five of the six *O*s in the case of the N room, and for all six in the case of the T room. Both the N and the T rooms were seen, at least to some degree, as having the shape of the L room, with only a single exception. In 9 of the 11 instances in which there was such carry-over, the *O*s said that they saw slightly less of the L-room appearance at the N or T rooms than they had at the L room, but that it was clear and unequivocal nevertheless. In the other two instances no reduction in perceived L-room shape was reported. In *O*s' reports the N and T rooms were described as looking as though the floor sloped down to the left, and in most cases one or more other L-room appearances such as ceiling sloping up to the left, back wall receding to the left, etc. were described.

A few illustrative excerpts from the protocols follow.

1. Observer C. S. viewing N room after Learning Session 3:

E. How does the floor look?
O. It looks as if it slants down to the left.
E. How does the ceiling look?
O. It slants up to the left some.
E. How about the windows?
O. The left window of the center panel looks larger and like it is slanting down to the left on the bottom.
E. Anything else you can tell me about the appearance of the room?
O. Well, I think it does give the appearance of being like the room that slants to the left, quite a lot.

2. G. S. viewing the T room after Learning Session 3:

E. How does the floor look?
O. The floor seems to tilt down toward the left.
E. How about the ceiling?
O. The ceiling tilts up and to the left.
E. How about the back wall?
O. It seems to slant quite a little bit. Back to the left.

[5] Perhaps it should be pointed out that this phenomenon is not at all analogous to classical figural aftereffect. The "transferred" configuration is not different, "compensatory," or "opposite," but is, to the degree that transfer occurs, like the learned configuration. In addition, visual fixation is not involved. Satiation or statistical theories of figural aftereffect should not be invoked as explanations for this phenomenon, as they clearly do not apply.

3. P. B. viewing the N room after Learning Session 4:

O. Well, the floor slopes down towards—down towards the bottom left-hand, you know, the far left-hand corner. Again the corner up at the right looks closer to me than the corner at the left and it looks shorter than the corner at the left. The ceiling looks like it's going up towards the left-hand side.

The results for Group NA were substantially the same as for Group A. Four of the six *O*s reported alterations in the appearance of the L room after Learning Session 1, and all six did so after Learning Session 2. Four of the six reported carry-over to the N room, and five of the six to the T room. However, these results plus a comparative examination of the protocols strongly suggest that the perceptual alteration was slightly less in speed and amount and that there was somewhat less carry-over to the N and T rooms than for Group A.

The replies to questions in the post-experimental interviews provide convincing evidence that *O*s experienced marked changes in visual space perception and reported what they saw. Here is an example that is perhaps more concise than most, but otherwise typical.

O. It progressively got more distorted as the experiment went on. I could tell some difference in the back wall and in the ceiling and floor. The ceiling seemed to be the slowest to come along, though. It seemed to be normal for quite a time there and then I began to see a difference—it began to slope, I think, upwards and to the left.

E. This is an actual change in how it looked?

O. That's right.

The evidence obtained in this experiment would seem to point to several rather clear-cut conclusions.

1. Marked alterations in visual space perception did occur as a consequence of the learning sessions. Every *O* did learn to see the L room in more nearly its true shape even though the stimulus configuration presented to his eye remained unchanged, thus confirming earlier findings.

2. This perceptual learning tends to increase with increasing amounts of practice in this situation.

3. The almost unanimous carry-over of the newly learned percept to both the N and the T rooms provides strong confirmation for the basic hypothesis that formative learning does occur. The *O*s did learn to see differently the basic visual configuration related to all three rooms, independently of give-away cues. We may infer that any properly constructed room based on the same reference room would have been seen the same way, regardless of its actual shape.

4. The evidence that the carry-over from the L room to the N and T rooms was not complete suggests that give-away cues functioning below the level of awareness also played a role in the learning. It seems likely that reorganizational learning did occur in the L room, but the cues on which it depended were not available in the N and T rooms. Only give-away cues to the N or the T shapes would be available in those rooms. Also, the protocols provide some evidence that the learning sessions resulted in a generalized increase in the saliency of give-away cues, not only for the L room, but for the other rooms as well. These factors would work against the carry-over of the newly learned percept.

5. The notion that gross overt action is necessary for such perceptual modification to occur is clearly wrong. The amount of action indulged in by Group NA was at a minimum, yet there was very little less perceptual learning than in Group A. A more adequate formulation of the role of action in perceptual learning is called for.

Discussion

It would be easy to speculate at length concerning the significance of these findings and conclusions for perceptual learn-

ing in general. However, it is felt that much more evidence is needed before any such ambitious project is undertaken. This discussion will be confined to a very few notions concerning reorganizational and formative learning, the role of action in perceptual learning, and the use of the distorted rooms in the controlled study of factors involved in learning in visual space perception.

It appears probable that both the reorganizational and formative processes are at work in almost any instance of perceptual learning; however, the formative process may be thought of as more fundamental in the sense that it is the one on which the expansion and development of our perception of our world depends. If we consider perceiving as acting, as responding, then we might say that reorganizational learning is simply the reshuffling of old previously learned responses; formative learning is the acquiring of new ones. Of course, the matter is not this simple, as there is no doubt an intimate and complex relationship between the two processes so that they affect one another differently under different circumstances. Perhaps if giveaway cues are sufficiently available to permit an adequate and stable perceptual reorganization to be arrived at easily and quickly, formative learning related to the basic configuration may not occur because it becomes unnecessary. On the other hand, when the reorganizational process occurs to some degree but not enough to provide such adequate reorganization, it probably facilitates formative learning through helping to stabilize the learned modification as it occurs.

Implicit in the above is the suggestion that formative perceptual learning is gradual and not subject to sudden shifts; for it, continuity is the rule. Reorganizational learning, on the other hand, probably proceeds in sudden shifts. A hitherto unutilized cue becomes effective, with a consequent sudden alteration in perception. Reorganizational learning *need not* give the impression of discontinuity since the shifts may be so small as to conceal their nature; but it *may* do so, as in the case of hidden pictures, the rotating trapezoid (1), etc.

Let us turn now to the problem of the role of action in perceptual learning. A more adequate formulation might be that some degree of perceptual modification occurs whenever certain aspects of the total process of perceiving are out of harmony with others in terms of carrying out some purpose, and this means both across modalities and within a single modality. Action merely serves the function of bringing the organism into contact with the cues that give rise to the disharmony; if the cues are made available in other ways, e.g., through the actions of others, the conditions necessary for perceptual learning will have been fulfilled. One might also add that probably the modification generally tends toward an organization which has more functional value in the sense of providing a better bet based on experience about "what happens next."

It should be pointed out that this reformulation does not appreciably diminish the importance of the role of action in perceptual learning. Apparently perceptual learning requires that a relevant set of sequential events be set in motion in such a way that successive impingements on the learning organism will be attended to and related over time to one another, to expectations, and to purposes. Only in rare circumstances, such as in an experiment, will these conditions even be approximated unless the learner is the operator. An operator other than the learner can provide the necessary sequence of conditions if he knows enough about what the learner perceives, what he expects, what he is attending to and will attend to, etc., but these factors will be known imperfectly in the first place and are continually changing in a unique way for the learner as the sequence of events proceeds. Thus, only when the learner is the operator can the conditions for perceptual alteration be maximized.

Discussion of the use of the distorted rooms as a means of investigating learning in visual space perception necessitates some preliminary remarks about previous work. Research involving postoperative cataract cases, the use of prism, aniseikonic, and inversion lenses, and the employment of unstructured or ambiguous pictures or objects would seem to have contributed most heavily to our knowl-

edge in this area. However, certain limitations in these investigations must be noted. The cataract and lens researches are, in general, characterized by lack of experimental control of the factors involved in the perceptual learning; the other experiments have generally involved rather sudden shifts from one fairly stable perceptual organization to another and made analysis of the processes involved in the change rather difficult. Perhaps a valuable supplement to these would be the use of the distorted rooms. They provide a clear-cut instance of learning in visual space perception gradual enough to be described or measured at intervals as it proceeds, and one in which such factors as amount of practice, kind of practice, motivation, advance knowledge, number and kind of conflicting visual cues available, etc. might be varied systematically and checked for effects. Also, they permit reorganizational and formative perceptual learning to be separated for experimental treatment.

Summary

The experiment reported in this paper was designed primarily to distinguish between two hypothesized perceptual learning processes, reorganizational and formative. A secondary purpose was to test the notion that gross overt physical action on the part of the learner is a necessary condition for perceptual learning.

Three small rooms, one cubical and two distorted, were used as basic apparatus. After experimentally demonstrating that the rooms are monocularly equivalent and seen as "normal" or cubical prior to learning, learning sessions were conducted at one of the distorted rooms. Six *Os* (the action group) tossed balls at a target and felt around the room with a wand while observing its interior monocularly; and six *Os* (the no-action group) observed the room's interior while *E*'s assistant tossed the balls and manipulated the wand.

For all 12 *Os*, the learning sessions resulted in reported alteration in appearance of the distorted room in which the practice occurred; 11 of the 12 *Os* reported that this alteration in appearance carried over to one or both of the other two rooms, but generally with a decrement. This transfer of the newly learned percept was as predicted and was interpreted as evidence that formative as well as reorganizational perceptual learning occurs. Some probable characteristics and relations of these two learning processes were discussed.

The positive results from the no-action group were clearly contrary to hypothesis, and a modified formulation concerning the role of action in perceptual learning was outlined.

Finally, it was pointed out that rooms of the kind used in this experiment provide a means for investigating many aspects of learning in visual space perception.

References

1. Ames, A., Jr. Visual perception and the rotating trapezoidal window. *Psychol. Monogr.*, 1951, **65**, No. 7 (Whole No. 324).
2. Cantril, H. *The "why" of man's experience.* New York: Macmillan, 1950.
3. Gibson, J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950.
4. Hoffman, E. L. The role of action in defining and changing visual perception. Unpublished doctor's dissertation, Princeton Univ., 1953.
5. Ittelson, W. H. The constancies in perceptual theory. *Psychol. Rev.*, 1951, **58**, 285–294.
6. Ittelson, W. H. *The Ames demonstrations in perception.* Princeton: Princeton Univ. Press, 1952.
7. Ittelson, W. H., & Kilpatrick, F. P. Equivalent configurations and the monocular and binocular distorted rooms. In F. P. Kilpatrick (Ed.), *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952. Pp. 41–55.
8. Kilpatrick, F. P. Assumptions and perception: three experiments. In F. P. Kilpatrick (Ed.), *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952. Pp. 153–173.
9. Kilpatrick, F. P. Statement of theory. In F. P. Kilpatrick (Ed.), *Human behavior from the transactional point of view.* Hanover, N. H.: Institute for Associated Research, 1952. Pp. 1–15.
10. Kilpatrick, F. P., & Ittelson, W. H. The size-distance invariance hypothesis. *Psychol. Rev.*, 1953, **60**, 223–231.
11. Koffka, K. *Principles of Gestalt psychology.* New York: Harcourt, Brace, 1935.
12. Köhler, W. *Gestalt psychology.* New York: Liveright, 1947.
13. Pratt, C. C. The role of past experience in visual perception. *J. Psychol.*, 1950, **30**, 85–107.

(Received for early publication January 7, 1954)

Journal of Experimental Psychology
Vol. 47, No. 5, 1954

THE EFFECT OF KINESTHETIC, VERBAL, AND VISUAL CUES ON THE ACQUISITION OF A LEVER-POSITIONING SKILL [1]

WILLIAM F. BATTIG

University of Wisconsin

Because of the many factors involved, the study of acquisition of a complex motor skill is difficult, unless the skill can be analyzed into its component parts in some way. One possible method of analysis is according to the kinds of cues or information provided during learning. The nature of these cues will vary widely from one motor skill situation to another. In order to get any general results through this approach, it may be necessary to investigate a wide variety of skills. However, a comparison of the relative roles of various types of cues within a single skill should be useful in that it will determine their relative effects on learning of the skill.

In the present experiment, an attempt was made to determine the role of several kinds of cues in the acquisition of a precise lever-positioning movement. This was done by comparing performance on the standard task following training in which certain cues were eliminated and others emphasized. The task required the movement of a joystick to a position which would light green lamps corresponding to red lamps located on a stimulus panel. Three types of cues are involved: (*a*) visual cues, provided primarily by the red and green lights on the panel; (*b*) kinesthetic cues, provided by making the actual movement of the stick; (*c*) verbal cues, provided by any sort of verbal description of the stimuli or required movements. This particular task appears to maximize visual cues, with a corresponding lack of emphasis on the other types.

This experiment is relevant to the following two hypotheses about motor learning:

a. The verbal hypothesis: Brogden (4) has hypothesized that the acquisition of a motor skill will be facilitated to the extent to which verbal descriptions of the stimulus, the response, or both, can be utilized. This hypothesis would predict that if verbal cues are added to the standard task, superior acquisition will result. However, if training procedures which emphasize verbal cues show a high degree of positive transfer to the standard task, this result would be interpreted as supporting the hypothesis.

b. The kinesthetic hypothesis: This was originally formulated and tested, with inconclusive results, in connection with a gunnery training device (14). It states that an emphasis on kinesthetic cues during training may result in a higher final level of performance on the skill. If it can be shown that training procedures emphasizing kinesthetic cues produce a high degree of positive transfer to the standard task, this hypothesis would gain some support.

[1] This research was supported in part by the United States Air Force under Contract Number AF 33(038)-23294 monitored by the 3301st Research and Development Group (Perceptual and Motor Skills Research Laboratory), Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas. Reproduction in whole or in part is permitted for any purpose of the United States Government. The author wishes to acknowledge the assistance of Dr. W. J. Brogden, under whose guidance this study was done.

Little previous work has been done with direct bearing on the present problem, but a number of studies in the area of motor transfer have some relevance. In a series of experiments, Gagné and associates have found positive transfer to a simple discriminatory motor reaction task from practice on components of the task (**11, 12**); from pencil and paper performance on a pictured representation of the task (**13**); and from verbal practice, both on the response (**1**), and on predifferentiation of the stimuli (**9**). Eckstrand, using a similar task, found positive transfer to making discriminatory responses to color from practice with verbal stimuli (**6**), and with different forms (**7**).

In addition, a number of studies on the Mashburn apparatus are somewhat related. Baker, Wylie, and Gagné (**2**) found evidence for both positive transfer and interfering effects when another lever-positioning movement was added to the standard task. A series of studies by Lewis and associates (**15, 16, 17**), on the effect of reversing the direction of control movement, have shown over-all negative transfer to the standard task, but also some evidence for facilitation.

Procedure

Apparatus.—The apparatus was the Complex Coordination Test, Model E (Mashburn), as described by Melton (**19**). Two major modifications were made for the purposes of the present experiment: (*a*) the rudder movement and corresponding row of lights were eliminated; (*b*) the self-stepping feature, by which each correct matching of the lights automatically presented a new stimulus setting, was also eliminated. This was necessary in order to equate the amount of training for all groups, as some of the training procedures could not be carried out under a self-paced situation. Instead, a fixed 20-sec. interval between successive stimulus settings was used.

The modified task can then be described as follows: The apparatus was placed in a booth, with a panel of lights set flush with the front end of the booth. The *S* was seated facing the light panel, with the joystick on the floor in front of him. The rudder control was present, but fixed so it would not move, and used only as a footrest.

The light panel consists of two parallel rows of red and green lights, and two parallel columns of red and green lights. There are 13 lights in each row and column. Movement of the stick from left to right controls the green lights in the horizontal row, movement forward and backward controls the vertical row. When the stick is in the neutral position, the center (7th) light in the row and column of green lights is lit.

The *S*'s task was as follows: The stimulus consisted of two illuminated red lights, one in the row and one in the column. The *S* would move the stick, starting from the neutral position, so as to light the green lights opposite each of the two red lights simultaneously as rapidly as possible. Each correct matching of the red lights constituted a trial. The score was the time required to make a correct match after the stimulus setting had been presented. After the completion of each trial, *S* returned the stick to the center position to await the next trial.

The *E* sat behind the booth and could not be seen by *S*. He was able to observe *S* and the light panel by means of a mirror arrangement. Observation of the brush contacts, located in the cabinet of the apparatus, provided a further means of checking *S*'s movements.

The red-light settings were controlled by a 40-position stepping switch operated manually by *E*. The apparatus was wired so that a series of 13 stimulus settings was repeated three times, with one setting repeated in the fortieth position, on each complete revolution of the stepping switch.[2]

Subjects.—Sixty male students from the University of Wisconsin, whose names were acquired from a file of students who had indicated an interest in serving as *S*s for pay, were used as *S*s.

Design.—The *S*s were divided into six groups of ten each. Five groups served for ten sessions, approximately 30 min. in length, every day, Monday through Friday, for two weeks. The sixth group served for only two sessions, spaced 11 days apart.

All groups performed the standard task on the first and last of the ten sessions, designated as the

[2] Many *S*s were able to recognize this order of presentation of the stimuli to some extent. At the conclusion of the experiment, all *S*s were run through the cycle and asked to anticipate which stimulus was coming next. There was considerable individual variability, and little or no relationship to the time scores. The quantitative data can be found in the thesis upon which this report is based (**3**), along with other data which are not presented here.

pretest and posttest, respectively. The groups were differentiated by their tasks during the middle eight sessions, the training period, as follows:

Group A (Standard practice): This group performed the standard task throughout the training period.

Group A' (No practice): This group did nothing during the training period.

Group B (Verbal stimulus): The red lights were turned off and covered by two boards on which were printed rows of numbers, one opposite each green light, which indicated the position in terms of number of lights away from the center. The words "back," "forward," "left," and "right" were also printed on the boards, indicating the direction of movement. The task for this group differed from the standard task only in that the stimulus consisted of two numbers read by *E* (i.e., "six right, three forward") rather than two red lights. This training procedure eliminates visual cues, in part, and emphasizes the use of verbal cues.

Group C (Verbal response): The red light settings were presented exactly as in the standard task, but instead of making any response with the stick, *S* simply described verbally how he would move the stick to reach the correct position, in terms of distance from the center and direction of movement. This procedure eliminates kinesthetic cues entirely, and also some visual cues, while placing emphasis on verbal cues.

Group D (Kinesthetic response): The red light settings were presented as in the standard task, but the green lights were turned off, and *S* attempted to move the stick to the correct position without the aid of the green lights. When *S* had reached what he judged to be the correct position, the green lights were turned on again so he could see how close he had come. This procedure eliminates visual cues, in part, and emphasizes kinesthetic cues.

Group E (Verbal stimulus-kinesthetic response): Neither the red nor green lights were used. The *S*'s task was to move the stick in response to verbal descriptions given by *E*. The descriptions consisted of the angular displacement of the stick from the left-right plane, and the direction and extent of movement (i.e., "move it about 6 in. to your left and away from you, at an angle of about 30°"). If the movement was incorrect at first, further description was provided until the correct position had been reached. This procedure virtually eliminates visual cues, emphasizing kinesthetic cues and, to some extent, verbal cues.

Groups A and A' are designated as control groups. The other four groups, which practiced on modifications of the standard task during the training period, will be referred to as experimental groups.

Instructions.—The standard instructions (19) were given at the beginning of the experiment, modified slightly because of changes in the apparatus and procedure. The instructions given to the experimental groups on the second day were worded, as much as possible, in the same way. At this time, all *S*s in the experimental groups were told that the purpose of the experiment was to find the effect of training-task practice upon subsequent performance on the standard task.

No instructions were given as to the method of matching the lights. There were some individual differences, but most *S*s used the successive method (lining up first one light and then the other) in preference to the simultaneous method.

A ready signal was given shortly before the presentation of each new stimulus. During the standard task (and the training procedure for Group B), *S* was told the time it had taken him to make the correct match after each trial.

General procedure.—Each matching of a red-light setting constituted a trial. All groups completed 60 trials in each daily session, divided into three blocks of 20 each, separated by a 1-min. rest period. Within each block, trials were given at regular 20-sec. intervals.

The red-light patterns were presented in the order wired into the stepping switch, as described previously. In each session, the cycle of 40 settings was run through one and one-half times. Each successive block on a given day started at the point where the previous one had left off, so that the third block of trials on each day was identical to the first. The starting point within the cycle was varied from session to session, but was the same for the pretest and posttest.

During a given session, the order of presentation of stimulus settings (or descriptions) was identical for all groups, so that the order and difficulty of trials were constant.

The experiment was run in five 2-week periods, during which 11, 11, 11, 9, and 8 *S*s, respectively, were run. A progressive matching procedure based on the pretest scores was adopted in determining the assignment of *S*s to groups. As a result, all groups (except Group A') were equated in terms of initial ability, being well matched in both means and variability. The proportion of *S*s assigned to each group during each 2-week period was approximately the same.

The *S*s in Group A' (the no-practice group) had to come for only two sessions, and therefore were assigned to this group prior to any determination of initial ability. As a result, their performance averaged about 10% lower than that for the other groups on the pretest scores, making it necessary to introduce special procedures in analysis of the data.

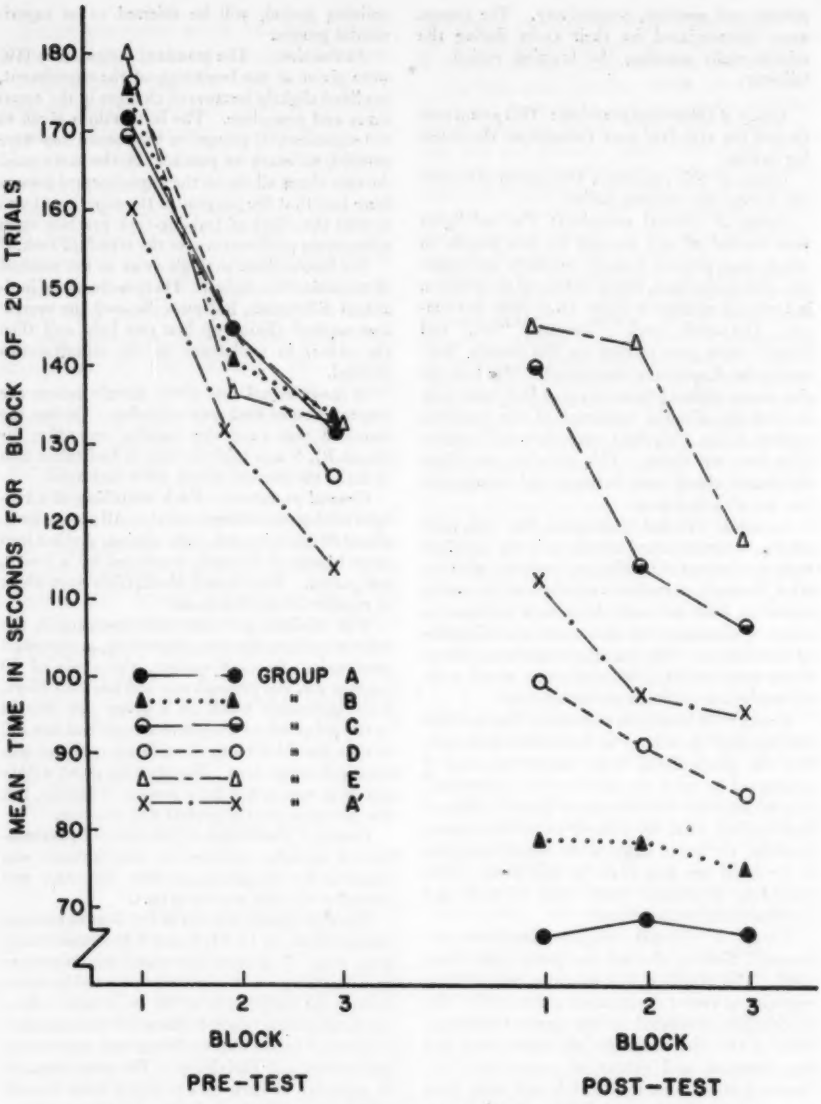


FIG. 1. Mean total time to complete each block of 20 trials on pretest and posttest

RESULTS

The mean total time in seconds required to match the 20 settings of each block of the pretest and posttest is given in Fig. 1 for each of the six groups.

Inability to match Group A′ (no practice) with the other groups in ini-

tial ability resulted in a 10% difference between the pretest scores of Group A′ and the other groups, as shown in Fig. 1. This difference is not statistically significant ($t = 1.69$, $.20 > p > .10$), but is large enough so that it should be taken into account.

Consequently, Group A′ has been compared separately with each of the other groups by means of t tests on the differences between the pretest and posttest scores.[3] This comparison showed that Groups A (standard practice), B (verbal stimulus), and D (kinesthetic response) showed significant improvement over no practice at the 5% confidence level (except for Block I for Group D). Groups C (verbal response) and E (verbal stimulus-kinesthetic response) were inferior to Group A′ (no practice), except for Block III for Group C. Only Group E on Block II differs significantly at the 5% level. This indicates that the training procedures for these two groups may have produced a decrement in performance on the standard task. Group A′ is excluded from further analysis.

Between-within analyses of variance showed significant differences between groups, yielding F ratios of 16.55 for Block I, 13.10 for Block II, 8.25 for Block III, and 13.79 for total posttest scores. $F = 2.58$ is significant at the 5% level.

A more complete analysis of variance showed significant differences between blocks of trials ($F = 30.96$) and a group by block interaction ($F = 6.94$). The values of F significant at the 5% level are 3.10 and 2.04, respectively. This shows a significant improvement with posttest practice which varies from group to group. Figure 1 shows the source of the interaction, a greater improvement by the groups highest on Block I. This can be attributed either to an initial interference effect between the training and standard tasks, or to the fact that Groups A (standard practice) and B (verbal stimulus) were approaching an asymptote of performance.

Duncan range tests (5) were employed in comparing individual pairs of groups. Groups A (standard practice) and B (verbal stimulus) differed significantly at the 5% level from Groups C (verbal response) and E (verbal stimulus-kinesthetic response), but not from each other. Groups C and E differed significantly only on Block II. Group D (kinesthetic response) did not differ from Group B, differed from Group A only for Block I and total, and differed from Groups C and E in all cases except Block II in the case of Group C.

Posttest scores for settings.—The posttest scores for the 13 red-light settings used were also analyzed by analysis of variance. Significant differences between settings ($F = 16.76$) and a settings-by-groups interaction ($F = 2.22$) were found. The values of F at the 5% level are 1.78 and 1.40, respectively. This indicates that the 13 settings varied in difficulty, and that the relative difficulty varied from group to group.

The order of presentation of the settings was the same throughout the experiment, which confounds order with difficulty. However, an empirical measure of difficulty was obtained by pooling the scores for all groups on

[3] At first glance, this may appear to be a violation in correct use of the multiple t test. However, Group A′ is not involved in the basic part of the experiment. It was included only to get a measure of the amount of improvement resulting from each of the experimental training procedures. Its role in this experiment is essentially the same as if four independent experiments had been performed, each to determine the degree of transfer resulting from one of the training procedures. If considered in this light, the use of the t test here appears to be justified.

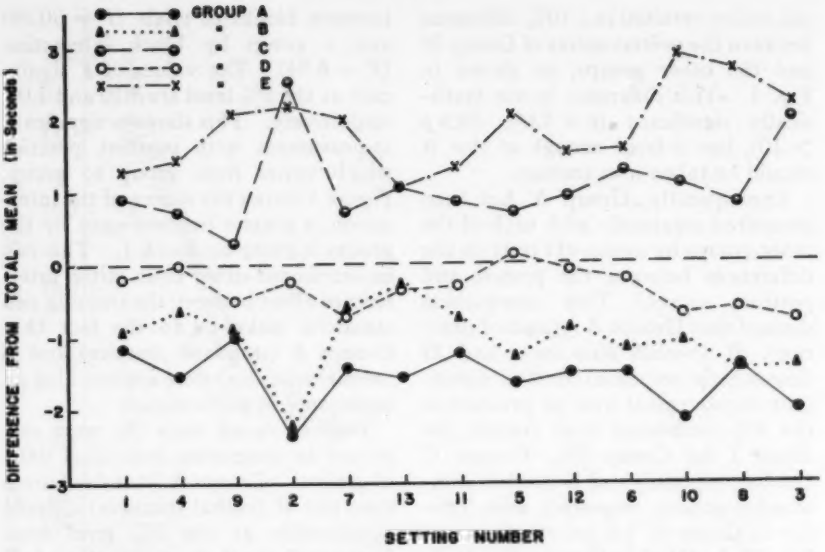


FIG. 2. Difference of each group from total mean for each stimulus setting. Settings are arranged on the abscissa in order of difficulty, as determined by pooled scores of all groups.

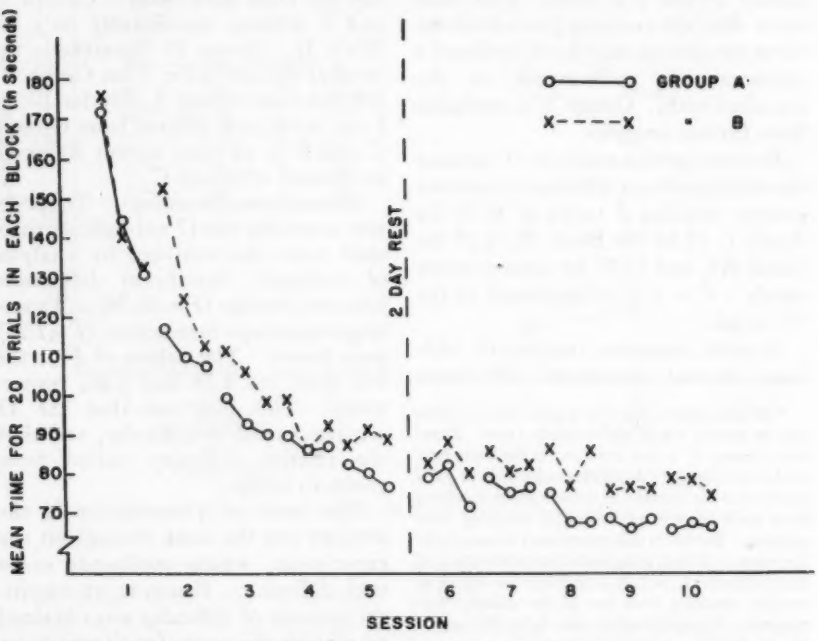


FIG. 3. Mean total time to complete each block of 20 trials for Groups A and B on each of the ten sessions

each setting. This revealed no consistent relation to the characteristics of the settings, either in terms of position or extent of movement.

Figure 2 represents an attempt to determine the source of the settings by groups interaction. The settings are ordered on the abscissa in terms of empirical difficulty. The curves represent the deviation of each group from the total mean for all groups for each setting. No apparent relation between group variation and empirical difficulty is shown.

Training scores.—The time scores for Groups A (standard practice) and B (verbal stimulus) during the training period (Sessions 2–9) are presented in Fig. 3, which shows a typical negatively accelerated learning curve, apparently approaching asymptotic performance.

The fact that stimulus presentation took longer for Group B (verbal stimulus) added a constant factor to the time scores, and prevents a direct comparison with Group A (standard practice). The increment of the Group B posttest scores over Session 9 indicates the presence of interference, probably greater than Fig. 3 shows, a conclusion which is supported by the marked difference between Groups A and B on Session 2.

Figure 4 gives training period scores for Group D (kinesthetic response), and shows the mean total number of lights off the correct position on each day (obtained by summing over the 60 trials) and the mean total number of correct settings on each day. These curves are rough reciprocals of each other, and indicate that learning was taking place under the Group D training procedure.

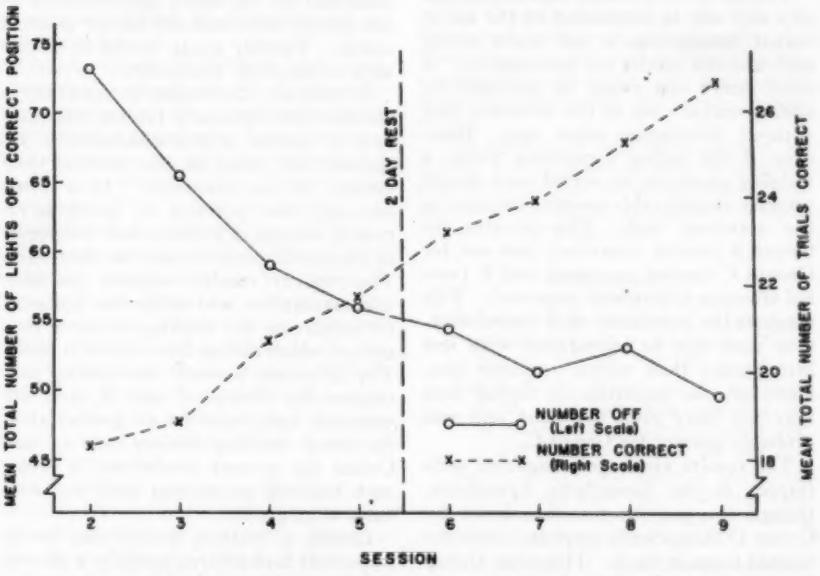


FIG. 4. Training data for Group D. Solid curve and left scale give mean total number of lights off correct position for each session. Dashed curve and right scale give number of trials correct for each session.

Discussion

The results indicate that none of the training procedures produces as high a level of performance as does practice on the standard task (Group A). Only Group B (verbal stimulus) was not significantly different, and its scores were consistently higher on all measures. This suggests the interpretation that visual cues are most important for this type of skill, and that elimination of visual cues impairs acquisition. However, the presence of interference between the standard task and training procedures restricts such an interpretation, as the present study confounded elimination of visual cues and interference effects. Possibly a standard task emphasizing kinesthetic cues would show poorer performance as a result of training emphasis on visual cues, and this would be attributable solely to interference. The use of a more direct method of comparing performance with and without visual cues will be necessary to resolve this question.

The verbal hypothesis that acquisition of a skill will be facilitated by the use of verbal descriptions is not under direct test, and the results are inconclusive. A more direct test would be provided by *adding* verbal cues to the standard task without eliminating other cues. However, if the verbal hypothesis holds, a training emphasis on verbal cues should produce considerable positive transfer to the standard task. This is true for Group B (verbal stimulus), but not for Groups C (verbal response) and E (verbal stimulus-kinesthetic response). This suggests the possibility that verbal stimulus cues can be introduced with less interference than verbal response cues. However, the emphasis on verbal cues may not have been the same and was probably greater for Group C.

The results are largely negative with respect to the kinesthetic hypothesis, though the positive transfer shown by Group D (kinesthetic response) provides limited support for it. However, Group E (verbal stimulus-kinesthetic response) showed no positive transfer, which is evidence against the hypothesis.

It appears that the greater the modification of the training task, the poorer the subsequent performance on the standard task, and that response modification produces more interference than stimulus modification. This coincides with results in verbal learning, where transfer varies as a function of task similarity, response dissimilarity having a greater negative effect (18).

Comparison of the posttest scores with the learning curve for Group A (standard practice) in Fig. 3 gives a rough quantitative measure of efficiency of the training procedures. Group B (verbal stimulus) corresponds to Group A on Sessions 6 and 7, Group D (kinesthetic response) is surpassed by Group A on Session 4, while Groups C (verbal response) and E (verbal stimulus-kinesthetic response) are surpassed by Group A on Session 2. Eight days of training for Groups C and E were evidently less effective than a day of rest for Group A.

A longer posttest might have eliminated differences between groups; this is indicated by the rapid improvement of the groups with initially higher posttest scores. Further study would be necessary to establish this point.

Of primary importance in this study is the fact that Groups C (verbal response) and E (verbal stimulus-kinesthetic response) did worse on the posttest than Group A′ (no practice). It is often assumed that practice on anything remotely related to a motor skill will result in improved performance on that skill. The present results seriously question this assumption and stress the necessity for validating any training device or procedure which differs from the skill itself. The difference between the training procedures for Groups C and E, and the standard task, may be no greater than for many training devices now in use. Unless the present results are in error, such training procedures may do more harm than good.

Length of training period may be an important factor here; possibly a shorter training period would show no negative transfer. This is a subject for future investigation. Loss of motivation is a

possible variable, but the learning curves for Groups A (standard practice), B (verbal stimulus), and D (kinesthetic response) show no evidence for a loss of motivation, and there is no indication that Groups C (verbal response) and E (verbal stimulus-kinesthetic response) should differ in this respect.

A possible explanation for the results of Group E is that these *S*s had trouble relating the training and standard tasks. They tended to sit forward and look down at the stick, a departure from normal posture. Also, they were trained in making diagonal movements, and many did not adhere to this method on the posttest. It has been shown that diagonal movements on this apparatus result in poorer performance early in learning than do separate movements directly in the two planes (19).

The present results conflict with previous studies in the area, especially those (1, 20, 21, 22) which have shown positive transfer from verbal practice to a motor task, and with Eckstrand's finding (6) that practice of the response to one type of stimulus shows positive transfer to another type.

Gagné, Baker, and Foster (10), in a theoretical analysis of the relation between motor skill similarity and transfer, conclude that complete or partial reversal is the only type of task alteration which leads to negative transfer. The training procedures for Groups C (verbal response) and E (verbal stimulus-kinesthetic response) do not involve any such reversals, except insofar as Group E made a diagonal response during training and responded separately in the two planes on the posttest. Since in the one case the response would be made to both stimuli in conjunction, and in the other case to the two stimuli separately, this might constitute what Gagné *et al.* call partial reversal.

It is difficult to relate this study to Gagné's theory, and it may be that the present experiment involves a situation not covered by the Gagné, Baker, and Foster paper. The tasks were very dissimilar, as Gagné used a simple discriminatory motor reaction in which each stimulus was associated with a distinct response. Also, the training period was much shorter than that of the present study. Nevertheless, the results strongly suggest that other types of modification besides reversal can produce negative transfer.

Summary

An attempt was made to determine the roles of kinesthetic, verbal, and visual cues in the acquisition of a lever-positioning skill by comparing performance following training procedures in which these cues were differentially emphasized. All *S*s performed the same task on the first and last of ten practice sessions, but six groups were differentiated on the basis of their task during the middle eight sessions.

A comparison of posttest scores showed that Group A (standard practice) was superior to all other groups. Of the experimental groups, Group B (verbal stimulus) was best, followed in order by Group D (kinesthetic response), Group C (verbal response), and Group E (verbal stimulus-kinesthetic response).

Groups A, B, and D showed significant improvement over a group with no training practice (Group A′), but Groups C and E did not.

It is concluded that training on any modification of the standard task produces performance inferior to that following standard practice. Long periods of practice on tasks closely related to the standard task may even produce negative transfer.

The presence of interference between the standard and training tasks makes it difficult to compare the roles of the various types of cues, but the results point to the importance of visual cues for this particular skill.

References

1. Baker, K. E., & Wylie, R. C. Transfer of verbal training to a motor task. *J. exp. Psychol.*, 1950, **40**, 632–638.
2. Baker, K. E., Wylie, R. C., & Gagné, R. M. The effects of an interfering task on the learning of a complex motor skill. *J. exp. Psychol.*, 1951, **41**, 247–260.
3. Battig, W. F. The effect of kinesthetic, verbal, and visual cues on the acquisition of a motor skill. Unpublished master's thesis, Univ. of Wisconsin, 1953.
4. Brogden, W. J. Unpublished memorandum, Univ. of Wisconsin, 1952.
5. Duncan, D. B. A significance test for differences between ranked treatments in an

analysis of variance. *Virginia J. Sci.*, 1951, **2**, 171–189.

6. Eckstrand, G. A. Response practice as a factor in transfer of training. Dayton, O.: U. S. Air Force, Air Materiel Command, Wright-Patterson Air Force Base, 1950. (*AF Tech. Rep.*, No. 6017.)
7. Eckstrand, G. A. Cue attention habits as a factor in training. Dayton, O.: Wright-Patterson Air Force Base, 1951. (*Tech. Rep.*, No. 6566.)
8. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
9. Gagné, R. M., & Baker, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.
10. Gagné, R. M., Baker, K. E., & Foster, H. On the relation between similarity and transfer of training in the learning of discriminative motor tasks. *Psychol. Rev.*, 1950, **57**, 67–79.
11. Gagné, R. M., Baker, K. E., & Foster, H. Transfer of discrimination training to a motor task. *J. exp. Psychol.*, 1950, **40**, 314–328.
12. Gagné, R. M., & Foster, H. Transfer of training from practice on components of a motor skill. *J. exp. Psychol.*, 1949, **39**, 47–68.
13. Gagné, R. M., & Foster, H. Transfer to a motor skill from practice on a pictured representation. *J. exp. Psychol.*, 1949, **39**, 342–354.
14. Gregg, L. W., Hornseth, J. P., Nagel, E. H., & Brogden, W. J. Kinesthetic and visual factors in sighting proficiency. Unpublished manuscript, Univ. of Wisconsin, 1951.
15. Lewis, D., McAllister, D. E., & Adams, J. A. Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, **41**, 247–260.
16. Lewis, D., & Shephard, A. H. Losses of skill in performing the standard Mashburn task arising from different levels of learning on the reversed task. Port Washington, L. I., N. Y.: U. S. Navy Special Devices Center, 1950. (*Tech. Rep.*-SDC 938-1-9.)
17. McAllister, D. E., & Lewis, D. Facilitation and interference in performance on the modified Mashburn apparatus: II. The effects of varying the amount of interpolated learning. *J. exp. Psychol.*, 1951, **41**, 356–363.
18. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* New York: Longmans, Green, 1952.
19. Melton, A. W. *Apparatus tests.* Washington: U. S. Government Printing Office, 1947. (*AAF Aviat. Psychol. Program Res. Rep.* No. 4.)
20. Perry, H. M. The relative efficiency of actual and "imaginary" practice in five selected tasks. *Arch. Psychol., N. Y.*, 1939, No. 243.
21. Sackett, R. S. The relationship between amount of symbolic rehearsal and retention of a maze habit. *J. gen. Psychol.*, 1935, **13**, 113–128.
22. Vandell, R. A., Davis, R. A., & Clugston, H. A. The function of mental practice in the acquisition of motor skill. *J. gen. Psychol.*, 1943, **29**, 243–250.

(Received for early publication November 7, 1953)